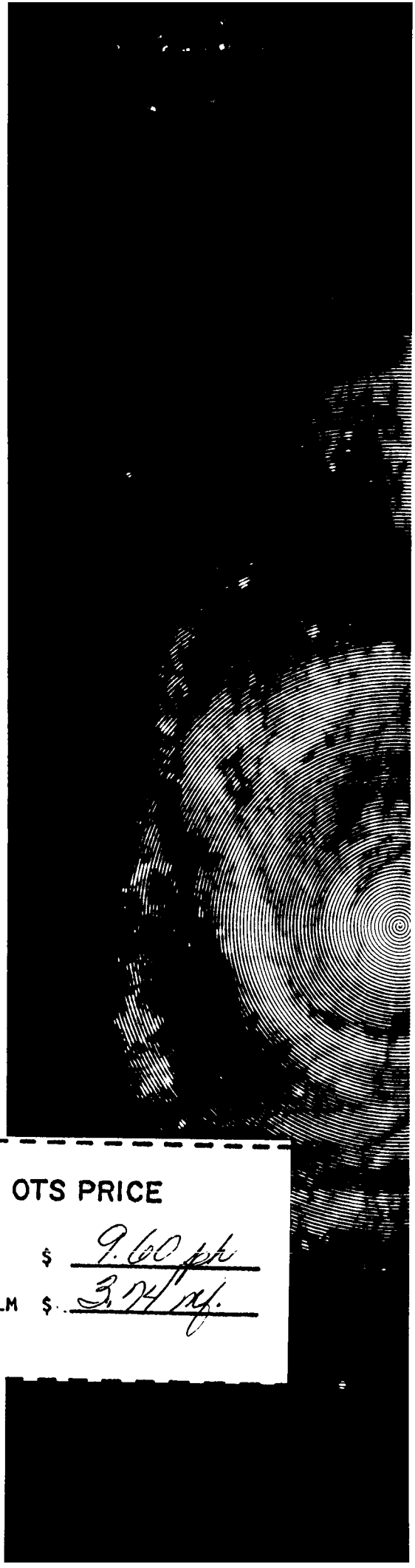


118 p.

Report 129-F

CLOUD PATTERN INTERPRETATION

Final Report
National Aeronautics and Space Administration
Contract NASw-609
August 1963

ASTROPOWER, INC.
SUBSIDIARY OF THE DOUGLAS AIRCRAFT COMPANY, INC.

2121 PAULARINO AVENUE, NEWPORT BEACH, CALIFORNIA

Report 129-F

CLOUD PATTERN INTERPRETATION

Final Report
National Aeronautics and Space Administration
Contract NASw-609
August 1963

NASA

cite only
(NASA CR-53778; Rept. 129-F) OTS: $--- see cover

R. D. Joseph, S. S. Viglione, and H. F. Wolf Aug. 1963 118p refs

ASTROPOWER, INCORPORATED
Newport Beach, California

03/6 502 Electronics Lab.

FOREWORD

This report was prepared by Astropower, Inc., Newport Beach, Calif., in fulfillment of NASA contract NASw-609. The work was administered by the Directorate of Electronics and Control, Office of Advanced Research and Technology, NASA Headquarters, Washington, D.C. Mr. H. Lowell was Project Engineer for this office.

The studies presented began on March 4, 1963. This document is the final report on the contract. Three status reports have been issued previously in the form of informal progress letters.

Work on the contract was performed by the Electronics Laboratory of the Research Division of Astropower. The chief contributors were R. D. Joseph, S. S. Viglione, and H. F. Wolf. The optical correlation measurements were performed by the Optical and Infrared Section of the Electromagnetic Radiation Laboratory of the Douglas Aircraft Co. in Santa Monica, Calif., under the direction of Mr. S. Spinak.

SUMMARY

19587

Using actual TIROS films and an optical correlation procedure, the feasibility of using a self-organized, parallel logic system to recognize cloud structures indicative of storms was investigated. The films were a sampling of cloud formations, some with and some without vortices, available from the plethora of data returned by the successful TIROS meteorological satellite system. This data was standardized to attempt to remove the variations due to film type and processing technique, and to concentrate on the only item of interest, the cloud structure. The optical correlation measurement was performed, and a computer routine was implemented to analyze this test data. The computer routines mechanized the equations of an approximate "forced learning" perceptron analysis which would enable the estimation of the number of logic units required in the parallel logic layer to achieve a specified recognition performance level.

The computer results showed that due to the intermixing of the pattern classes, caused partly by the inaccuracies of the optical measuring procedure and partly by the complexity of the patterns themselves, the forced learning perceptron analysis achieves only a limited effectiveness of recognition (65-70%). The required logic layer to asymptotically approach this performance level consists of only 50 logic units. An increase in the number of logic units does not significantly improve the system performance.

Means are discussed for increasing the effectiveness of the decision process by implementing a closed-loop learning routine (iterative design) developed concurrently with this program. Utilization of this technique in a digital simulation made (in conjunction with a discriminant analysis procedure to define the logic unit inputs) to analyze a larger sample of data in digital form should lead to the design specification of a parallel logic recognition system capable of making the required cloud pattern interpretation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1.0 INTRODUCTION

On April 1, 1960, TIROS I was launched into orbit to prove the feasibility of meteorological satellites. Five succeeding TIROS satellites have since been successfully operated to transmit television pictures and infrared scans of the earth's cloud cover. The quality of the data obtained and the reliability with which it was delivered were so high that, commencing with TIROS III, the TIROS series can be considered as an operational weather satellite system, although it did, up to TIROS VI, allow daily coverage of only about 18% of the earth's surface. Of the seven TIROS satellites still to come, some will be placed in highly elliptical orbits and some in synchronous orbits to provide more coverage.

The NIMBUS project is the follow-on to TIROS in the NASA meteorological satellite program. The NIMBUS concept specifically calls for worldwide coverage on a regular basis. The NIMBUS system will observe every point on the earth's surface twice a day, once in daylight and once at night.

In parallel with the development of the actual satellites and supporting ground systems, NASA and the National Weather Satellite Center have also started to develop automatic means of processing the cloud cover information to increase the speed of information flow from the Command and Data Acquisition Station to other weather centrals, and to facilitate the work of the satellite meteorologists. The first step in this direction was to devise a computer routine that automatically plots a latitude and longitude grid upon an overlay to be used with the TIROS cloud cover picture. This grid is drawn from orbit tracking and telemetry information before the actual cloud picture arrives. For the advanced TIROS and NIMBUS systems, more advanced data processing is contemplated. The geographical grid, as well as landmarks, will be placed on the cloud pictures in analog form. Also under consideration is the use of a special facsimile encoder which would directly convert the picture information to electronic form suitable for facsimile transmission. In parallel, the video signal will be digitalized and placed in proper format for entry into a computer.* The computer, utilizing the telemetry data and orbital elements previously received, will rectify the video data (pictures are taken from different view angles) from an entire orbit and produce a mosaic in map form. Picture elements

*Johnson, D. S., Hall, W. F., and Bristor, C. L., "NIMBUS Data in Operational Meteorology," Astronautics and Aerospace Engineering, Apr 1963.

will be summarized into rectified grid squares. Single-digit descriptors will express the percentage of cloud cover and the average brightness. These digital mosaics will be presented in analog form to a manual processing unit, where meteorologists, utilizing the individual high resolution pictures, will superimpose meteorological interpretations such as atmospheric stability, wind flow, storm centers, fronts, etc. The analysts will utilize projected images with the interpretative information automatically fed back to the computer during the analysis phase. The final output, a merger of computer and manual analysis, will be transmitted to other weather centrals.

This highly automated processing technique is necessary because the amount of data expected from NIMBUS is too large to be processed manually. A feeling for the dimensions involved may be gained by considering that NIMBUS television cameras can deliver 96 pictures with a total of 5×10^8 bits of data per orbit.* As noted, all routine functions of the processing task will be accomplished by computer. The task which cannot be achieved automatically is the meteorological interpretation of cloud features and patterns. This work defies automatic handling, primarily because mechanisms capable of recognizing and interpreting pictorial information of the complexity of cloud photographs are not available. If a machine were developed that could recognize and interpret such cloud patterns as storm centers, cloud "streets," and fronts, it is conceivable that future weather forecasting systems could be completely automated.

In order to study the feasibility of such an automatic cloud pattern recognition system, NASA headquarters, Directorate of Electronics and Data Processing, awarded a study contract to the Electronics Laboratory of Astropower, Inc. The system originally proposed by Astropower is a self-organized, parallel logic, distributed memory recognition device coupled to a digital conditional probability computer and an operator control console. A block diagram of the entire system is shown in Figure 1-1. The system is such that it operates completely automatically; it is possible, however, for an operator to monitor the machine and override the automatic control.

The heart of the system consists of the parallel logic interpreter and the conditional probability computer. Figure 1-2 is a logic flow chart of the

*Schneebaum, M. I. and Stampfl, R. A., "Data Storage for Meteorological Satellites," Astronautics and Aerospace Engineering, Apr 1963.

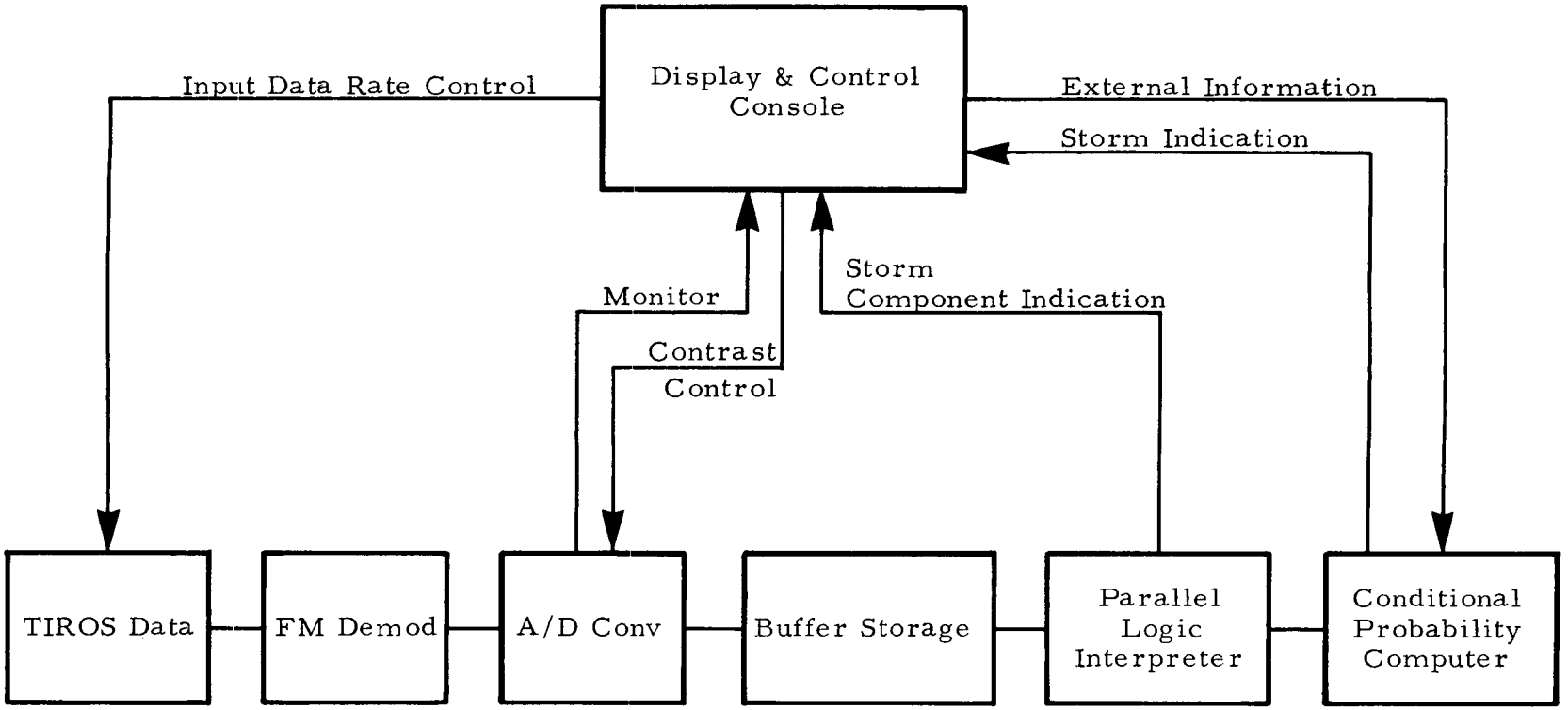


Figure 1-1. Block Diagram of Computer Macrostructure

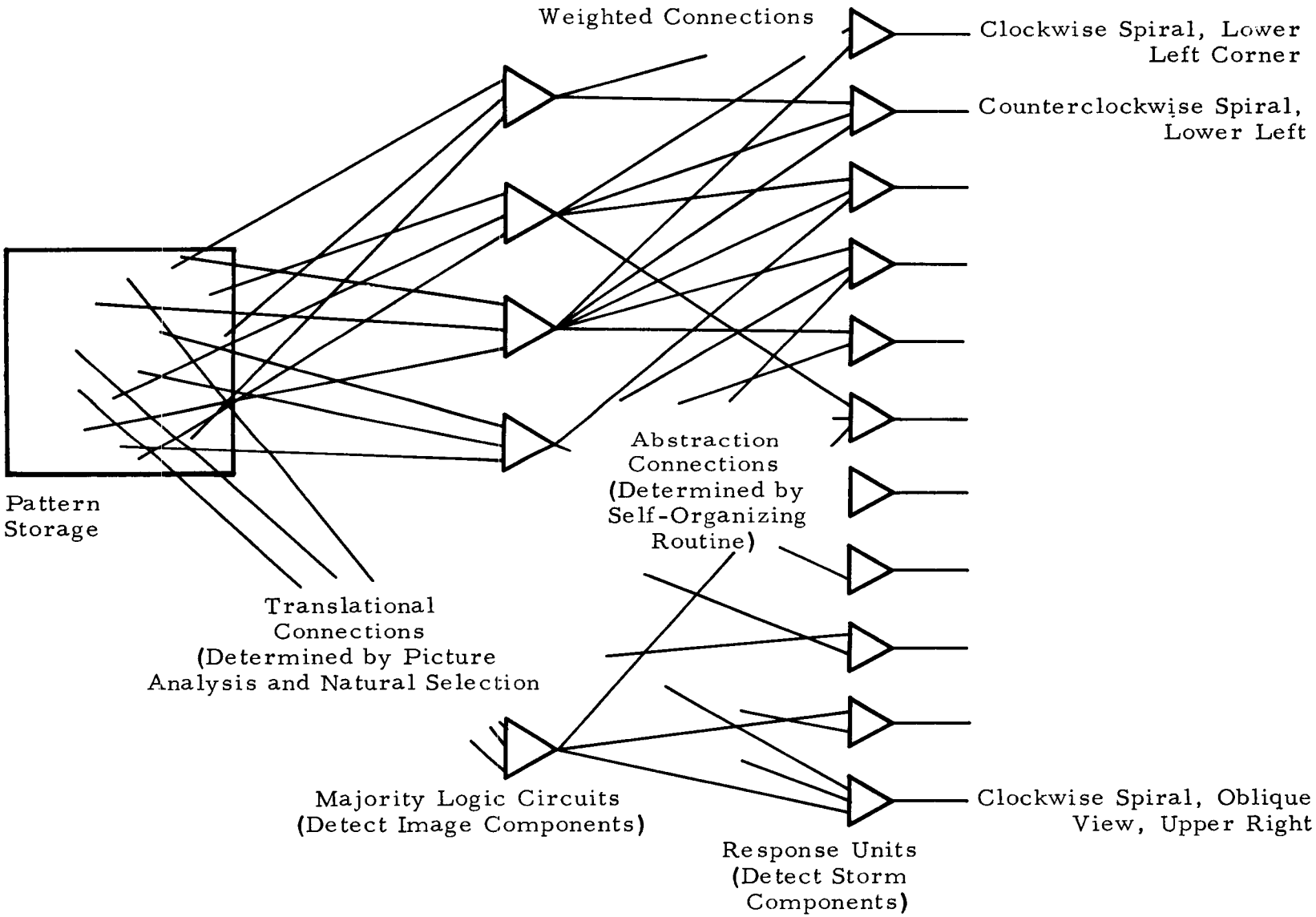


Figure 1-2. Block Diagram of Parallel Logic Intepreter

parallel logic interpreter. The input connections for each majority logic unit sample the cloud pattern in storage. The logic units compute majority logic propositions concerning the cloud pattern, the particular proposition being determined by the input connections. Thus each majority logic unit becomes a property filter, recognizing a specific property of the incoming pattern. The input connections themselves represent a sampling of a much larger list of property filters. This list may be selected completely at random, determined entirely from the known properties of the patterns, or compiled by combinations of these approaches - the method of selecting the property filter list depends upon how well the bases for pattern classification are understood. The self-organizing routine determines the utility of the various property filters included in the machine.

The conditional probability computer accomplishes the task of a table-lookup. The range of the output of each storm component indicator is partitioned into a discrete number of intervals. The output state of the parallel logic interpreter is thus digitized. The conditional probability computer has stored in it a table providing the conditional porbability of a storm for each of the finite output states. Any continuous function of the storm component indicators can be approximated to any degree of accuracy by choosing fine enough partitions.

It was suggested that the orderly development of such a system could be accomplished in three phases:

Phase I - Feasibility Study

a. Investigate the complexity of the problem to arrive at a reasonable estimate of the size of the parallel logic interpreter in terms of the number of logic units required for the recognition of vortex* structures.

b. Study the system problems associated with the construction of a feasibility model.

c. Investigate the availability of suitable off-the-shelf hardware for the construction of a feasibility model.

*A vortex is a spiral-shaped cloud pattern indicating the center of a cyclonic storm, such as a hurricane or typhoon.

Phase II - Design and Construction of Breadboard Model

a. Design and construct feasibility model for the recognition of vortex structures.

b. Provide final proof of the ability of a parallel logic-conditional probability machine to analyze detailed cloud pictures.

c. Provide design details for a complete interpreter.

Phase III - Design and Construct Complete Self-Organized Conditional Probability Interpreter

The work performed under contract NASw-609 was directed to the Phase I study to prove the feasibility of this approach to cloud pattern interpretation by estimating the size of a parallel logic interpreter for the recognition of vortices. Theoretical optimization studies were also undertaken to reduce the complexity of the logic layer by increasing the efficiency of each logic unit and by optimizing the logic unit input connections and threshold setting. Finally, the hardware problems connected with the parallel logic approach to cloud pattern interpretation were investigated.

This report describes the work accomplished during the contract period. A review of self-organized pattern recognition machines is given in Section 2.0. The general machine organization, the learning rules applied during the self-organizing period, and several procedures for reducing the size of machine required for a given performance are discussed. Section 3.0 gives the mathematical derivations upon which the estimation of machine size is based. Section 4.0 treats in detail the selection and preparation of the cloud photographs, the design of the optical correlator, and the optical measurements. Section 5.0 discusses the computer program that was written to evaluate the optical data and gives the results of the program. A short discussion of the hardware implementation problems is presented in Section 6.0. Conclusions and recomendations are summarized in Section 7.0.

2.0 SELF-ORGANIZING, PARALLEL LOGIC SYSTEMS

Pattern classification must often be based on incompletely specified criteria empirically derived from "typical examples" of the pattern classes. While this method is characteristic of humans, it has only recently been adapted to automata. Machines which use this method are called self-organizing systems.

2.1 Organization of Parallel Logic Systems

In the devices to be considered, the patterns are to be classified on the basis of a discrete sampling of the pattern data. The samples obtained from a particular pattern class may be conceptualized as a set of coordinates which represents the pattern as a point in a k-dimensional signal space. The design of an automatic pattern classification device then is largely the specification of a partition on this space such that the cells of the partition may be identified with unique classes of patterns. The design constraints must include the maximum probability of error acceptable after the specification of the partition, and the limitations on the cost necessary to automate the process of specifying the classification regions (the decision process).

When the patterns associated with unique classes are concentrated in well-defined and widely separated regions of the signal space, simple and highly accurate decision mechanisms may be constructed. In more difficult designs, a useful approach is to employ simple decision devices (to achieve low cost) in fairly elaborate structures (to achieve the required partition complexity). Such a structure is diagrammed in Figure 2-1.

The "S" units are the sensory or input units to the device. In the case of pictorial data, a sensory unit is associated with a picture element and generates a signal proportional to the brightness of that element. The sensory field is connected to the first-layer logic units through weighted connections to translate the incoming signal to a form more suitable for recognition. The input signal may be translated through several layers of logic units, but eventually there is an inevitable "necking down" of the data - referred to as abstraction.

The performance of a parallel logic system may be characterized in terms of the response units. The response units considered here are

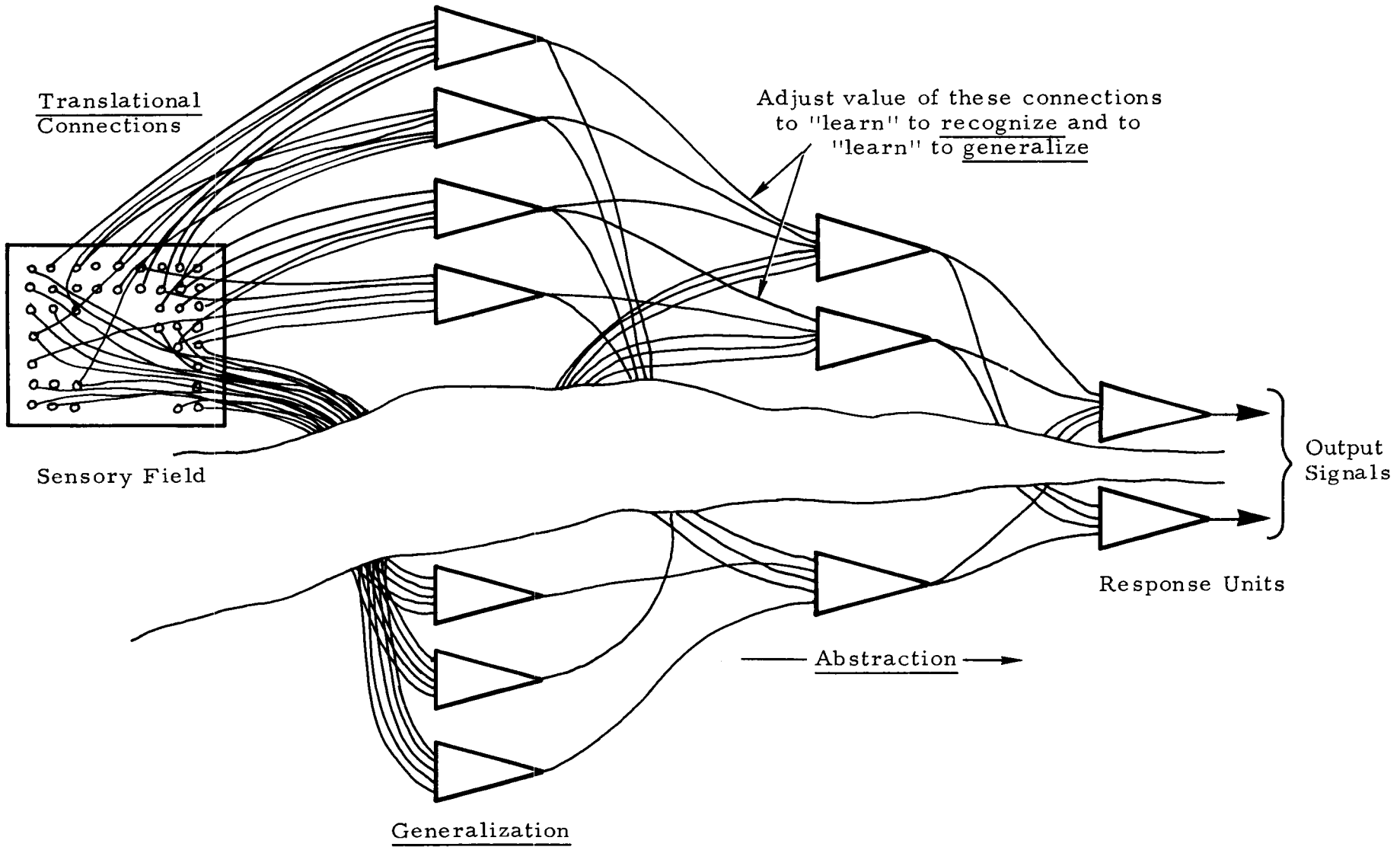


Figure 2-1. Generalized Parallel Logic, Distributed Memory System

two-state devices required to be in one state when a particular pattern is present in the signal displayed to the sensor units, and in a second state when this pattern is not present.

Initially there is little organization in these machines. However, by showing a sequence of signals to the machine and varying the weights of the connections between logic layers, or between logic and response layers, the machine's ability to classify patterns is improved.

A significant feature of these machines is their generalization capability. Since the classification of a pattern is made on a statistical basis, many logic units are active for each correct classification. The existence of a slightly different property in a signal of the same class as that upon which the machine was trained activates a large protion of the appropriate logic units, enabling correct classification despite minor variations in the pattern.

The decision units in Figure 2-1 are linear logic units*. A linear logic unit partitions the subspace defined by the sources of its inputs by passing a single hyperplane through the subspace, assigning a classification of "one" to patterns falling on one side of the hyperplane and "zero" to patterns falling on the other side. The linear logic unit mechanism is diagrammed in Figure 2-2. The rule governing the behavior of this unit is as follows: if the weighted sum of the input $\sum_i w_i e_i$ exceeds a threshold, θ, then an output, e_o, is generated and the unit is said to be active.

$$e_o = \begin{cases} 1 \text{ if } \sum_i w_i e_i \geq \theta \\ 0 \text{ otherwise} \end{cases}$$

The simplest organization which would give the machine a capacity to perform significant tasks contains two layers of logic (Figure 2-3)**. The sensory field again provides the input, and a linear logic unit serves as a report-out unit. Between the two is inserted a layer of linear

---

* Winder, R. O., "Threshold Logic in Artificial Intelligence," Artificial Intelligence Sessions, IEEE Winter General Meeting, Jan 1963.

** It should be noted that many pattern recognition devices are based on this structure - for example, the designs of Bledsoe and Browning, Gambs, Kamensky, Rosenblatt, Widrow, and the Astropower Decision Filter.

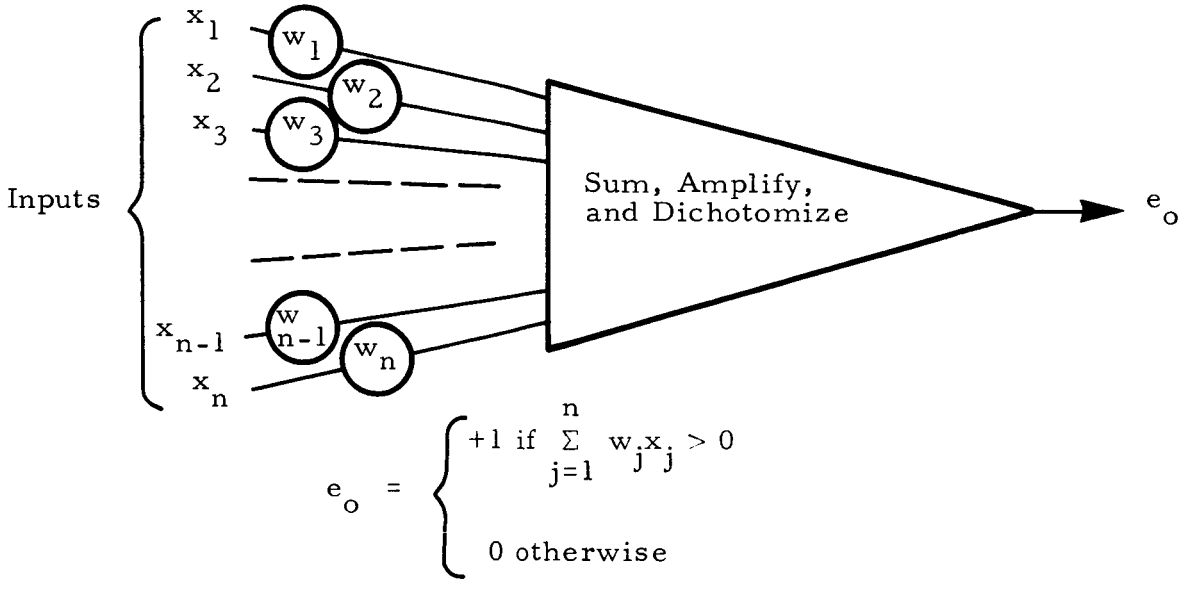


Figure 2-2. Linear Logic Unit Mechanism

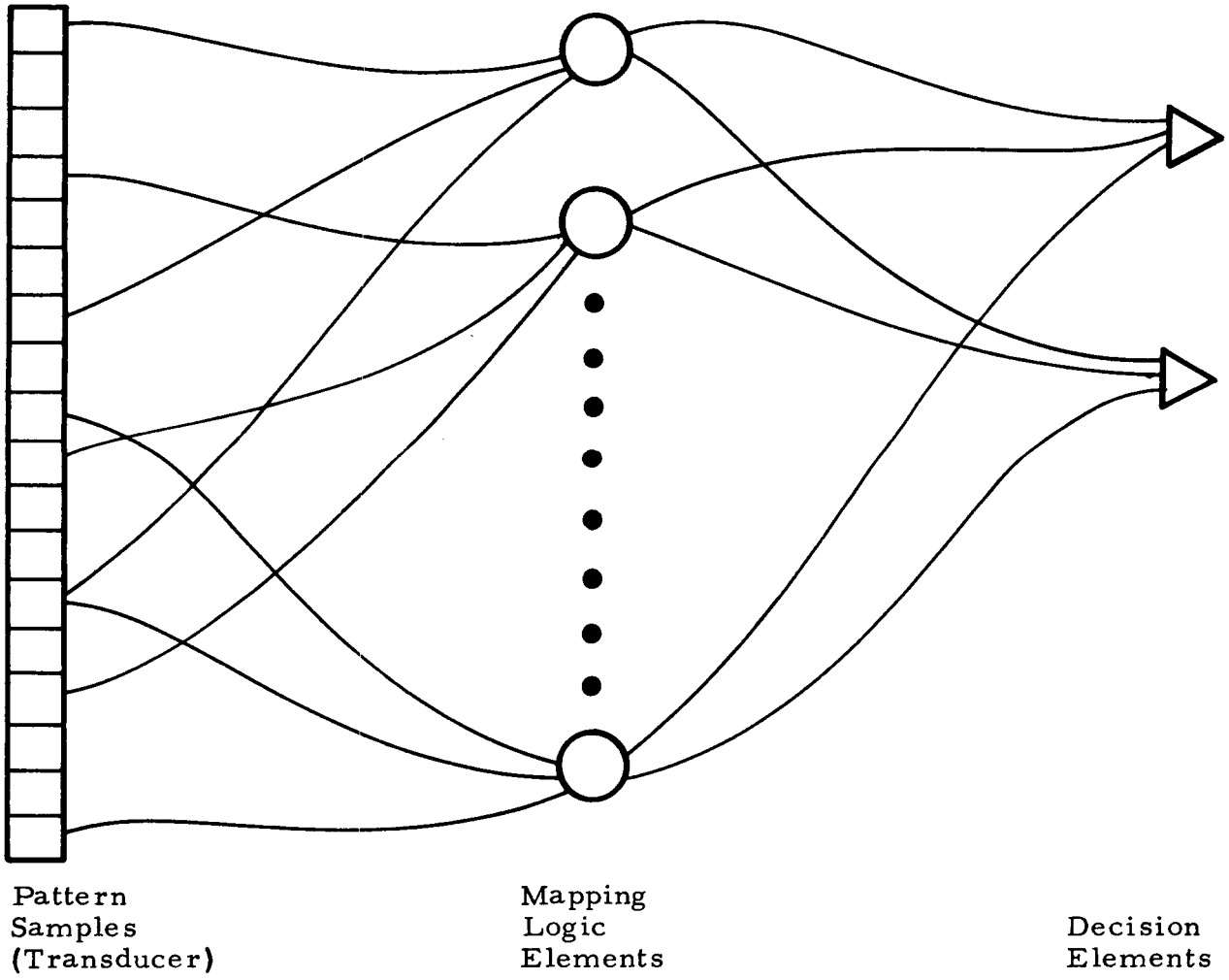


Figure 2-3. Pattern Recognition Network Organization

logic units. All of the connections shown are weighted. The connections from sensory points to logic units are many-to-many, and from logic units to report-out units are all-to-all.

2.2 Self-Organizing Algorithms

After selecting this organization as the network structure, a large number of network parameters remain to be specified. These parameters include the number of input connections, and the source points of these inputs for each logic unit; the weights of all connections from the sensory field to the logic units, and from the logic units to the decision elements; and the thresholds of all logic units and decision elements. A rule for automatically determining the values of these parameters based on a sample of classified patterns is called a self-organizing algorithm.

Consider one particular algorithm called "forced learning." When a logic unit is supplied with a set of input connections, it becomes a property filter,* determining whether or not the input patterns possess the particular property. A list of properties is established, and a sampling of these is selected for the recognition machine. The self-organizing procedure then determines the utility of the various property filters and the efficiency of the property list in making the required classification. The learning rule is such that it increases the weights of the connections from active units to the report-out unit when a pattern of interest is present, and reduces the connection weight of active units when no target or pattern of interest is present. With such a procedure, reliable indicators of target presence (that is, units which measure properties possessed by patterns of interest) will tend to carry large positive weights. Units which are reliable indicators that a target is not present, and hence measure properties possessed primarily by nontargets, will tend to carry large negative weights. Units which are less reliable indicators of the presence or absence of patterns of interest will carry positive or negative weights of lesser magnitude.

---

* Singer, T. R., "Model for a Size Invariant Pattern Recognition System," paper presented at the Bionics Symposium, Dayton, O., Sept 1960; Martin, T.B., and Talavage, J. F., "Application of Neural Logic to Speech Analysis and Recognition", paper presented at the Bionics Symposium, Dayton, O., 1963.

The list of property filters from which the logic unit input connections are selected is quite significant. The strucutre of this list determines the task for which the machine is suitable - that is, if the recognition machine is to perform properly, it must have the foundations upon which to function. Many techniques are available for generating this property list. One method is to manually build into the list each known parameter of the patterns to be classified. Such a procedure is time-consuming and tedious, and above all, it may not lead to a machine capable of correct classification if the discrimination clues are subtle and not known to the designer. One approach to this problem is to incorporate those properties known to be useful, then generate additional properties randomly and let the self-organizing procedure select the discrimination criteria. Such a procedure requires an extensive property list and may be accommodated by performing the self-organization in a simulated mode on a large scale general purpose digital computer.

The forced learning procedure is readily implemented in a computer simulation. It is well documented mathematically and gives rise to the "Theorem of Statistical Separability."*

For this theorem to hold, it is required that the property list emphasize the differences between targets and nontargets, and the similarities between these classes. The theorem makes it possible to compute whether or not the property list meets this requirement and to establish the required size of the logic layer for a specified performance level. It is the basis for the present program and is discussed thoroughly in Section 3.1.

Other algorithms have been developed for the design of efficient networks by increasing the effectiveness and the efficiency of the logic structure. One powerful tool is the application of a statistical discriminant analysis to assign weights to logic unit inputs from the sensory field. This produces a property list which is closely matched to the particular problem.

---

* Joseph, R. D., "Contributions to Perceptron Theory," Ph.D. Thesis, Cornell University, Sept 1961.

In addition to the means for assigning input connection weight, algorithms may be classified by whether they are open-loop or closed-loop processes, and by whether they are selective or not. A selective process is one in which the logic units generated are evaluated to determine their suitability for inclusion in the final network; in a nonselective process, all logic units generated are included in the network. Selective algorithms are best implemented by computer simulation of the self-organizing process. They produce much more efficient networks then do nonselective algorithms.

In an open-loop algorithm, the weight of a connection from a logic unit to the decision elements depends upon the on/off state of the logic unit for each pattern in the classified sample, but not upon the state of the decision elements. In closed-loop processes, the state of the decision elements is also significant in the determination of these weights. Closed-loop algorithms yield more complex self-organizing routines, but are effective for a broader range of problems. Table 2-1 gives examples of each category of algorithms.

TABLE 2-1

SELF-ORGANIZING ALGORITHMS

	Open-Loop	Closed-Loop	
Nonselective	Forced learning	Error correction	Easier implementation
Selective	Natural selection	Iterative design	Greater efficiency
	Easier implementation	Greater effectiveness	

One of the earliest successful algorithms was the perceptron forced learning rule briefly discussed previously. A detailed analysis of this algorithm is presented in the following section. With this rule, the weight assigned to the output connection of a logic unit is proportional to the difference between the number of sample patterns of one class which

turn the unit "on" and the number of sample patterns of the other class activating the unit. Forced learning is the only one of the algorithms of Table 2-1 for which adequate analytical tools exist. For this reason, it played a central role in this investigation.

Subsequent sections describe the natural selection (3.2.1) and iterative design algorithms (3.2.2). Discussions of the error correction algorithm may be found in the literature.*

* Block, H. D., "The Perceptron: A Model for Brain Functioning," Reviews of Modern Physics, Vol 34, No. 1, Jan 1962; Rosenblatt, E., Principles of Neurodynamics, Spartan Press, 1962.

3.0 ESTIMATION OF MACHINE SIZE

A two-stage estimation process for predicting the size of the parallel logic network required for the recognition of vortex patterns was undertaken. First, the number of logic units needed in a perceptron-type network was to be estimated. Using analytical tools developed at Astropower, it was possible to estimate this number without actually designing the perceptron. The work performed under this contract was primarily directed at accomplishing this task. Second, this estimate was to be modified to reflect the improvement in logic unit efficiency obtainable with more advanced self-organizing algorithms. No direct means exist for estimating the number of logic units required when these algorithms are used, and only limited empirical evidence is available to show the improvements achieved.

3.1 Perceptron-Type Networks

In this section, an exact statistical analysis of perceptron type networks (see Figure 2-1) is presented. An approximate computing form is then derived. Finally, means for estimating the necessary parameters utilizing optical correlation are described.

3.1.1 Exact Perceptron Analysis

The problem to be considered is the classification of a large number of multidimensional vectors into two categories. The data in each signal point are to be processed by a large number of logic units simultaneously; each logic unit determines the veracity of some linear logic proposition when applied to the vector under examination. As a result of this detailed examination of the individual vector, that vector receives a classification.

Suppose that the sensory field under consideration contains n points, so that the weighted input connections to a given logic unit can be represented as an n-dimensional vector. The j-th component of this vector represents the weight of the input connection from the j-th sensory point. The absence of a connection is represented by a zero.

The connection vectors for the logic units are selected randomly as follows. A distribution is assigned to the space of all possible input connection vectors (n-space), and the connection vector for each logic unit is selected independently according to this distribution. The set of connection vectors assigned non-zero probability is called the property list. In

this case, the property list was generated by assuming that the weight of the input connections to each logic unit is "A," the sum of squares of these weights is "B" (the numbers A and B being the parameters to be optimized), and that the points of origin of each input connection are selected independently, according to a uniform distribution over the entire sensory field.

Let C_i denote the input connection vector for the i-th logic unit, and S^k the input pattern expressed as an n-dimensional vector. The output of the i-th logic unit when the k-th pattern is shown, which will be denoted by δ_i^k, can be expressed in terms of the inner product of the connection vector and the signal vector

$$\delta_i^k = \begin{cases} 1 \text{ if } C_i \cdot S^k - \theta \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

where θ is the threshold of the unit. The outputs of the logic units pass through variably weighted connections to serve as inputs to the report-out units.

It is assumed that there are M input patterns, each to be grouped into one of two classes. With each pattern, there is an associated number δ which indicates the desired classification of the pattern. Thus

$$\delta^k = \begin{cases} 1 \text{ if } S^k \text{ is to be classified in the positive class} \\ -1 \text{ if } S^k \text{ is to be classified in the negative class} \end{cases}$$

A sequence of patterns is selected. The patterns are drawn from the population of patterns to be classified. Any individual pattern may appear in this sequence several times, or it need not appear at all. The number of appearances of the pattern S^j in the sequence is denoted by n^j. Following the selection of the connection vectors, this sequence of patterns is shown to the machine, during which the weights of the variable connections are modified.

The connections from the logic units to the report-out unit have variable weights. These weights are initially set to zero and then varied according to the following rule: if the i-th logic unit is activated when pattern S^k is shown during the adaptive sequence, then δ^k is added to the weight of the connection from the i-th unit to the report-out unit. Otherwise, the weight is unchanged. This can be restated as: the weight of the connection from the i-th unit to the report-out unit is changed by $\delta^k \delta_i^k$ each time S^k is shown during training. This is known as the "Forced Learning" rule.

Thus the weight of the i-th connection will be

$$w_i = \sum_{k=1}^{M} n^k \delta^k \delta_i^k$$

after the adaptive sequence. Consequently, if the input to the report-out unit when S^t is shown is denoted by β^t, then

$$\beta^t = \sum_{i=1}^{N} w_i \delta_i^t = \sum_{i=1}^{N} \sum_{k=1}^{M} n^k \delta^k \delta_i^k \delta_i^t$$

The report-out unit is said to have made a correct decision if

$$\beta^t \gtrless \theta_o$$

for

$$\delta^t = \pm 1 \text{ respectively.}$$

Consider the factors influencing β^t, and hence the classification of S^t:

δ^k = Previously determined classification for S^k

δ_i^k = Determined by connection vector of i-th logic unit

n^k = Choice of frequency of S^k in the adaptive sequence

Note that it follows that with a particular choice of network, adaptive sequence, and classification scheme, the decision on the test pattern, S^t, is completely deterministic.

When it is presumed that a random selection of connection vectors has been made, then β^t may be analyzed as a random variable. This assumption is the basis of the work which follows.

Developing β^t as the sum of N <u>independent</u> random variables permits an evaluation of the variance:

$$\sigma^2(\beta^t) = \sum_{i=1}^{N} \sigma^2(\beta_i^t) = \sum_{i=1}^{N} \left\{ E(\beta_i^t)^2 - \left[E(\beta_i^t) \right]^2 \right\}$$

$$\beta_i^t = \sum_{k=1}^{M} n^k \delta^k \delta_i^k \delta_i^t$$

$$(\beta_i^t)^2 = \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k \delta_i^j \delta_i^k \delta_i^t$$

Defining the following symbols:

$$Q_{jt} = E\left\{\delta_i^j \delta_i^t\right\} =$$ Probability of selecting a connection vector such that the i-th associative unit is active for both S^j and S^t. Note, under the assumption of random selection of connection vectors, that Q_{jt} is independent of i.

$$Q_{jkt} = E\left\{\delta_i^j \delta_i^k \delta_i^t\right\} =$$ Probability of selecting a connection vector such that the i-th associative unit is active for S^j, S^k, and S^t. Comment above applies.

One may write:

$$E\left(\frac{1}{N} \beta^t\right) = \sum_{j=1}^{M} n^j \delta^j Q_{jt} \qquad (1)$$

and

$$\sigma^2(\beta_i^t) = \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k \left\{Q_{jkt} - Q_{jt} Q_{kt}\right\}$$

Hence:

$$\sigma^2\left(\frac{1}{N} \beta^t\right) = \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k \left\{Q_{jkt} - Q_{jt} Q_{kt}\right\} \qquad (2)$$

Once a threshold for the response unit is selected, the mean and variance of β^t may be used in the Tschebycheff inequality to provide an absolute bound on the error probability. It may also be assumed that β^t is approximately normally distributed (since it is the sum of a large number of independent, well-behaved variables) to provide more reasonable estimates of the error rate.

One may also consider the test pattern S^t as having been selected at random either from the class of positive patterns (cloud formations

containing vortices) or from the negative class. In this case, the variance of the input to the response unit should contain a component due to the selection of the test pattern. One obtains

$$E\left(\frac{1}{N}\beta_{\pm}\right) = \frac{1}{M_{\pm}} \sum_{t} \sum_{j=1}^{M} n^j \delta^j Q_{jt} \tag{3}$$

and

$$\sigma^2\left(\frac{1}{N}\beta_{\pm}\right) = \frac{1}{NM_{\pm}} \sum_{t} \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k (Q_{jkt} - Q_{jt} Q_{kt})$$

$$+ \frac{1}{M_{\pm}} \sum_{t} \sum_{j=1}^{M} \sum_{k=1}^{M} n^j n^k \delta^j \delta^k Q_{jt} Q_{kt}$$

$$- \left[\frac{1}{M_{\pm}} \sum_{t} \sum_{k} n^k \delta^k Q_{kt}\right]^2 \tag{4}$$

where the summations on t are restricted to patterns in the appropriate class. Note that the variance has two components: the first is the average of the variances of β^t and may be made arbitrarily small by increasing the number of logic units; the second is the variance of $E(\beta^t)$ over the selection of a test pattern, and is not affected by a change in the number of logic units. The second component limits the performance which may be achieved with a perceptron.

3.1.2 Approximate Perceptron Analysis

The computation of the exact variance of β^t is very difficult because the numbers Q_{jkt} are quite numerous and are not readily available. Q_{jkt} may be obtained by a process similar to the one described in the following section for Q_{jt}, but a vastly expanded measurement program would be required. An approximate method of estimating the variance of β^t is thus desired.

An analysis of the terms contributing to the variance in Equation 2 indicates that the most significant ones occur when j = k. The approximation

$$Q_{jkt} \cong \frac{Q_{jk} Q_{kt}}{Q_k} \tag{5}$$

is exact when j = k. Substituting Equation 5 into 2,

$$\sigma^2\left(\frac{1}{N}\beta^t\right) \cong \frac{1}{N}\sum_j\sum_k n^j n^k \delta^j\delta^k\left[\frac{Q_{jk}\,Q_{kt}}{Q_k} - Q_{jt}\,Q_{kt}\right] \tag{6}$$

Considering the test pattern to be randomly selected,

$$E\left(\frac{1}{N}\beta_\pm\right) = M_\pm \sum_t \sum_{k=1}^{M} n^k \delta^k Q_{kt} \tag{7}$$

and

$$\begin{aligned}\sigma^2\left(\frac{1}{N}\beta_\pm\right) = {} & \frac{1}{NM_\pm}\sum_t\sum_{j=1}^{M}\sum_{k=1}^{M} n^j n^k \delta^j\delta^k Q_{kt}\left(\frac{Q_{jk}}{Q_k} - Q_{jt}\right) \\ & + \frac{1}{M_\pm}\sum_t\sum_{j=1}^{M}\sum_{k=1}^{M} n^j n^k \delta^j\delta^k Q_{jt}\,Q_{kt} \\ & - \left[\frac{1}{M_\pm}\sum_t\sum_k n^k \delta^k Q_{kt}\right]^2\end{aligned} \tag{8}$$

where the summations on t are restricted to patterns in the positive class or negative class, as appropriate. The expressions in Equations 7 and 8 may be rewritten as

$$E\left(\frac{1}{N}\beta_+\right) = \frac{1}{M_+}\sum_t\left[M_+\,\tilde{Q}_{t+} - M_-\,\tilde{Q}_{t-}\right] \tag{9a}$$

$$E\left(\frac{1}{N}\beta_-\right) = \frac{1}{M_-}\sum_t\left[M_+\,\tilde{Q}_{t+} - M_-\,\tilde{Q}_{t-}\right] \tag{9b}$$

and

$$\begin{aligned}\sigma^2\left(\frac{1}{N}\beta_+\right) = {} & \frac{1}{NM_+}\left[\sum_{k=1}^{M} n^k \delta^k M_+\,Q_{k+}\left(\frac{M_+\,\tilde{Q}_{k+} - M_-\,\tilde{Q}_{k-}}{Q_k}\right)\right. \\ & \left. - \sum_t\left(M_+\,\tilde{Q}_{t+} - M_-\,\tilde{Q}_{t-}\right)^2\right] \\ & + \frac{1}{M_+}\sum_t\left(M_+\,\tilde{Q}_{t+} - M_-\,\tilde{Q}_{t-}\right)^2 \\ & - \left[\frac{1}{M_+}\sum_t\left(M_+\,\tilde{Q}_{t+} - M_-\,\tilde{Q}_{t-}\right)\right]^2\end{aligned} \tag{10a}$$

$$\sigma^2\left(\frac{1}{N}\beta_-\right) = \frac{1}{NM_-}\left[\sum_{k=1}^{M} n^k \delta^k M_- \tilde{Q}_{k-}\left(\frac{M_+ \tilde{Q}_{k+} - M_- \tilde{Q}_{k-}}{Q_k}\right) - \sum_t \left(M_+ \tilde{Q}_{t+} - M_- \tilde{Q}_{t-}\right)^2\right] + \frac{1}{M_-}\sum_t \left(M_+ \tilde{Q}_{t+} - M_- \tilde{Q}_{t-}\right)^2 - \left[\frac{1}{M_-}\sum_t \left(M_+ \tilde{Q}_{t+} - M_- \tilde{Q}_{t-}\right)\right]^2 \qquad (10b)$$

where the summation on t is again restricted to patterns of the appropriate class. The symbols $\tilde{Q}_{k+}$ and $\tilde{Q}_{k-}$ are used to indicate the average values of Q_{jt} for t in the positive class and negative class respectively.

The forms of Equations 9 and 10 were used in the feasibility study. To test the validity of the approximation of Equation 5, some further approximations (variances of $\tilde{Q}_{k+}$ and $\tilde{Q}_{k-}$ are small compared to their means squared) were made to yield

$$\frac{1}{N}\frac{E^2(\beta_+)}{\sigma^2(\beta_+)} = \frac{\tilde{Q}_+ (\tilde{Q}_{++} - \tilde{Q}_{+-})}{\tilde{Q}_{++} + \tilde{Q}_{+-}} \qquad (11)$$

where $\tilde{Q}_{++}$ is the value of $\tilde{Q}_{k+}$ averaged over all patterns, S^k, in the positive class; $\tilde{Q}_{+-}$ is the average of $\tilde{Q}_{k-}$ over the positive class; and $\tilde{Q}_+$ is the average of $\tilde{Q}_k$. The important parameter estimated by Equation 11 defines the performance when the threshold of the response unit is set to zero.

The simplified expression for $1/N\,(E/\sigma)^2$ given in Equation 11 has been checked for accuracy in several examples. Results of these checks are presented in Tables 3-1 through 3-3. Rosenblatt* provides tables of Q_{jt} for completely random connections, various sizes of patterns, pattern overlap, and numbers of connections. In the first two examples (Tables 3-1 and 3-2), these are used to estimate Q_{++} and Q_+. Joseph** provides exact answers against which the approximation is verified. In the third example

---

* Rosenblatt, F., "Tables of the Q-Functions for Two Perceptron Models," Cornell Aeronautical Laboratory Report VG-1196-G6 (1960).

** Joseph, R. D., "Contributions to Perceptron Theory," Cornell Aeronautical Laboratory Report VG-1196-G7 (1960).

(Tables 3-3), the estimates of Q_{++} and Q_{+-}, and the data which yield very accurate approximations for verification, were provided by a simulation routine.

The first two examples consider horizontal and vertical bars on a 20 x 20 input mosaic. The bars were 4 x 20, and to keep the pattern size uniform, those portions of the bars which were translated off an edge of the mosaic were allowed to reappear at the opposite edge. The first example considered the task of classifying all horizontal bars as positive and all vertical bars as negative. The second example placed every alternate bar, horizontal and vertical, in the positive class, and the remaining set in the negative class. The third example given is the Astropower Decision Filter*.

Table 3-1 considers the problem of dichotomizing vertical and horizontal bars. In Table 3-2 the problem is more difficult. Alternate vertical bars are in the positive class, along with alternate horizontal bars. Note in both cases that the estimate obtained by applying the approximation is always within a factor of 1.4 of the exact result. The approximation in these cases overestimates the number of logic units required for a given performance level.

The approximate formula was also verified against the Astropower Decision Filter. Although this problem is too complex to permit the use of an exact solution for comparison purposes, there are powerful approximation methods which utilize a digital computer simulation to yield a nearly exact result. The comparison is shown in Table 3-3.

3.1.3 Estimation Technique

This section describes how the approximations of Equations 9 and 10 of the preceding section may be combined to estimate the number of logic units required to achieve a given performance level, and the means by which the required parameters may be estimated using optical correlations of sample patterns.

The input to the response unit is a random variable due to the random nature of the selection of the logic units, and the random selection of the test pattern. The conditional mean and variance of the input given that the test pattern is a vortex pattern, and the conditional mean and variance

*Joseph, R. D., Kelly, P. M., and Viglione, S. S., "An Optical Decision Filter," Proc. IEEE (Aug 1963).

TABLE 3-1

Q-TABLE FOR HORIZONTAL (+) AND VERTICAL (-) BARS ON 20 x 20 RETINA

Number of Input Connections Per Logic Unit with Weight of		$\Sigma\tilde{Q}_{++}$	$\Sigma\tilde{Q}_{+-}$	$\tilde{Q}_{+}$	$\frac{1}{N}(E/\sigma)^2$		Exact Approx Error Factor
+1	-1						
X	Y				Approx	Exact	
3	3	0.14892	0.06340	0.05632	0.02269	0.02592	1.1
3	7	0.05408	0.01220	0.02475	0.01564	0.01523	1.0
4	4	0.25511	0.14600	0.08544	0.02324	0.02708	1.2
4	6	0.15396	0.06840	0.05848	0.02250	0.02473	1.1
5	5	0.37198	0.24740	0.11120	0.02237	0.02579	1.2

TABLE 3-2

Q-TABLE FOR ALTERNATE HORIZONTAL (+-) AND VERTICAL (+-) BARS ON 20 x 20 RETINA

Number of Input Connections Per Logic Unit with Weight of		$\Sigma\tilde{Q}_{++}$	$\Sigma\tilde{Q}_{+-}$	$\tilde{Q}_{+}$	$\frac{1}{N}(E/\sigma)^2$		Exact Approx Error Factor
+1	-1						
X	Y				Approx	Exact	
3	3	0.11408	0.09824	0.05632	0.00420	0.00578	1.4
3	7	0.03807	0.02821	0.02475	0.00368	0.00529	1.4
4	4	0.21222	0.18889	0.08544	0.00497	0.00646	1.3
4	6	0.12093	0.10143	0.05848	0.00513	0.00690	1.3
5	5	0.32457	0.29481	0.11120	0.00534	0.00664	1.2

TABLE 3-3

VALUES OF $1/N\,(E/\sigma)^2$ FOR THE ASTROPOWER DECISION FILTER

	Simulation Results	Use of Equation 11
$\frac{1}{N}(E/\sigma)^2$ for Sphere/Cube	0.150	0.145
$\frac{1}{N}(E/\sigma)^2$ for Pyramid/Ellipsoid	0.083	0.080

given that the pattern does not contain a vortex, may be computed using Equations 9 and 10. For a given test pattern, the input to the response unit is approximately normally distributed, since it is the sum of a large number of independent variables — the outputs of the logic units. If the number of patterns is large, and if the means within a class are nearly normally distributed, this approximation may be extended to a randomly selected test pattern. Given a threshold for the response unit, the conditional means and variances may be used to compute the false alarm and missed target rates.

The threshold of the response unit may be adjusted to give the desired ratio of false alarms to missed targets. Assuming that equal probabilities of each type of error are desired (as was done in this study), the threshold may be placed s standard deviations from the mean of each distribution, where s is given by:

$$s = \frac{E\left(\frac{1}{N}\beta_+\right) - E\left(\frac{1}{N}\beta_-\right)}{\sigma\left(\frac{1}{N}\beta_+\right) + \sigma\left(\frac{1}{N}\beta_-\right)} \qquad (12)$$

Note that the denominator of Equation 12 is a function of N, the number of logic units. As N grows large, s approaches a limiting value. If this limiting separation of the classes is greater than the minimum acceptable level, the number of units required to achieve the minimum acceptable separation may be computed.

The evaluation of Equation 12 is dependent on the evaluation of the quantities $\tilde{Q}_{t+}$ and $\tilde{Q}_{t-}$ for various values of t. This may be accomplished by evaluating Q_{jt} for all values of the subscript j, or by sampling the possible range of this index, as was done in this study. A method is thus required for estimating Q_{jt}, the probability that a logic unit will be activated by pattern S^j and by pattern S^t.

3.1.4 Estimation of Q_{jt} Using Optical Correlation

Let $F_{jt}(u, v)$ denote the bivariate distribution function of the transparencies S^j and S^t. The contribution of a given input connection to the total input to the logic unit may be regarded as being equal in magnitude to the transparency of the pattern at the origin point of the connection, times the weight of that connection. If the origins of the connections are chosen at random,

independently and according to a uniform distribution, the following analysis applies.

Denote by X_{ji} the input to the i-th connection when S_j is shown. X_{ji} and X_{ti} have the joint distribution function $F_{jt}(X_{ji}, X_{ti})$ - the same distribution function as the transparencies. The joint distribution of the total inputs X_j and X_t is given by the convolution of connection density functions, since the origin points are selected independently. Denote this distribution by $G_{jt}(X_j, X_t)$.

Distinguishing those parameters pertaining to the distribution F_{jt} by tildas (e.g., $\tilde{m}_j$) we have

$$m_j = A\,\tilde{m}_j \qquad \sigma_{jt} = B\,\tilde{\sigma}_{jt}$$

$$m_t = A\,\tilde{m}_t \qquad \sigma_t^2 = B\,\tilde{\sigma}_t^2$$

$$\sigma_j^2 = B\,\tilde{\sigma}_j^2$$

Thus it is necessary to estimate $\tilde{m}_j$, $\tilde{m}_t$, $\tilde{\sigma}_t^2$, and $\tilde{\sigma}_{jt}$ using optical measurements.

$$\tilde{m}_j = \int_{u,v} u\, d F_{jt}(u,v) = \int_{u} u\, d F_j(u)$$

Hence $\tilde{m}_j$ is given by the total light passed by the S^j transparency, normalized by dividing by the total light passed by an empty frame. $\tilde{m}_t$ is obtained similarly.

$$\tilde{\sigma}_j^2 = \int_{u,v} u^2\, d F_{jt}(u,v) - \tilde{m}_j^2$$

$\tilde{\sigma}_j^2$ is obtained by measuring the total light passed by two S^j transparencies exactly superimposed, normalizing as above, and subtracting $\tilde{m}_j^2$.

$$\tilde{\sigma}_{jt} = \int_{u,v} u v\, d F_{jt}(u,v) - \tilde{m}_j\, \tilde{m}_t$$

Thus $\tilde{\sigma}_{jt}$ is obtained by subtracting $\tilde{m}_j\, \tilde{m}_t$ from the normalized quantity of light passed by superimposing the S^j and S^t transparencies. Note that the correlation coefficient ρ is given by $\tilde{\sigma}_{jt}/\tilde{\sigma}_j\, \tilde{\sigma}_t$ for both F_{jt} and G_{jt}.

The calculation of Q_{jt} is greatly simplified by the assumption that the total input to a logic unit when a pattern is shown is approximately normally distributed. The pair (X_j, X_t) denoting the total input for patterns S_j and S_t, respectively, then would have a bivariate normal distribution.

The derivation of m_j, m_t, σ_t^2, and ρ is in no way dependent on the normality assumptions; the assumption only insures the sufficiency of this set of parameters.

The normality assumption is, then, that

$$dG_{jt}(X_j, X_t) \sim \frac{1}{2\pi\,\sigma_j\sigma_t\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(X_j-m_j)^2}{(\sigma_j)^2} - 2\rho\frac{(X_j-m_j)(X_t-m_t)}{(\sigma_j)(\sigma_t)} + \frac{(X_t-m_t)^2}{(\sigma_t)^2}\right]} dX_j dX_t$$

Given the normal approximation, and the requisite parameters, the next problem is to obtain Q_{jt}.

$$Q_{jt} = \frac{1}{2\pi\,\sigma_j\sigma_t\sqrt{1-\rho^2}} \int_{\theta}^{\infty}\int_{\theta}^{\infty} e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\frac{(X_j-m_j)^2}{(\sigma_j)^2} - 2\rho\frac{(X_j-m_j)(X_t-m_t)}{(\sigma_j)(\sigma_t)} + \frac{(X_t-m_t)^2}{(\sigma_t)^2}\right]} dX_j dX_t$$

By substitution, and letting $a = \frac{\theta-m_j}{\sigma_j}$ and $b = \frac{\theta-m_t}{\sigma_t}$

$$Q_{jt} = \int_{b}^{\infty}\int_{a}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[y^2 - 2\rho yz + z^2\right]} dydz$$

or

$$Q_{jt} = \frac{1}{2\pi} \int_{\frac{a+b}{\sqrt{2(1-\rho)}}}^{\infty} \int_{\frac{b\sqrt{2} - x\sqrt{1+\rho}}{\sqrt{1-\rho}}}^{\frac{x\sqrt{1+\rho} - b\sqrt{2}}{\sqrt{1-\rho}}} e^{-\frac{1}{2}(x^2+y^2)} dydx$$

The derivation of the computing form will be accomplished using polar coordinates. The derivation depends upon the location of the origin

with respect to the lines L_1: $y+cx = bd$ and L_2: $y-cx = -ad$. Figure 3-1 defines the significant regions. Each sector will be considered in turn, and then the boundaries will be considered.

(Note: $c = \sqrt{\frac{1+\rho}{1-\rho}}$, $d = \sqrt{\frac{2}{1-\rho}}$)

A necessary and sufficient condition that the origin falls in Sector I is that $a > 0$ and $b < 0$. It is readily shown that

$$Q_{jt} = \frac{1}{2\pi} \int_{-\tan^{-1}c}^{\tan^{-1}\frac{c(b-a)}{b+a}} \int_{\frac{bd}{c\cos\varphi + \sin\varphi}}^{\infty} e^{-\frac{\gamma^2}{2}} \gamma d\gamma d\varphi$$

$$+ \frac{1}{2\pi} \int_{\tan^{-1}\frac{c(b-a)}{b+a}}^{\tan^{-1}c} \int_{\frac{ad}{c\cos\varphi - \sin\varphi}}^{\infty} e^{-\frac{\gamma^2}{2}} \gamma d\gamma d\varphi$$

The integration on γ is performed to obtain

$$Q_{jt} = \frac{1}{2\pi} \int_{-\tan^{-1}c}^{\tan^{-1}\frac{c(b-a)}{a+b}} e^{-\frac{b^2d^2}{2}\left(\frac{1}{c\cos\varphi + \sin\varphi}\right)^2} d\varphi$$

$$+ \frac{1}{2\pi} \int_{\tan^{-1}\frac{c(b-a)}{a+b}}^{\tan^{-1}c} e^{-\frac{a^2d^2}{2}\left(\frac{1}{c\cos\varphi - \sin\varphi}\right)^2} d\varphi$$

With suitable manipulations, this becomes

$$Q_{jt} = \frac{1}{2\pi} \int_{-\tan^{-1}c}^{\tan^{-1}\frac{c(b-a)}{a+b}} e^{-\frac{b^2}{2}\sec^2\left(\varphi - \tan^{-1}\frac{1}{c}\right)} d\varphi$$

$$+ \frac{1}{2\pi} \int_{\tan^{-1}\frac{c(b-a)}{a+b}}^{\tan^{-1}c} e^{-\frac{a^2}{2}\sec^2\left(\varphi + \tan^{-1}\frac{1}{c}\right)} d\varphi \qquad (13)$$

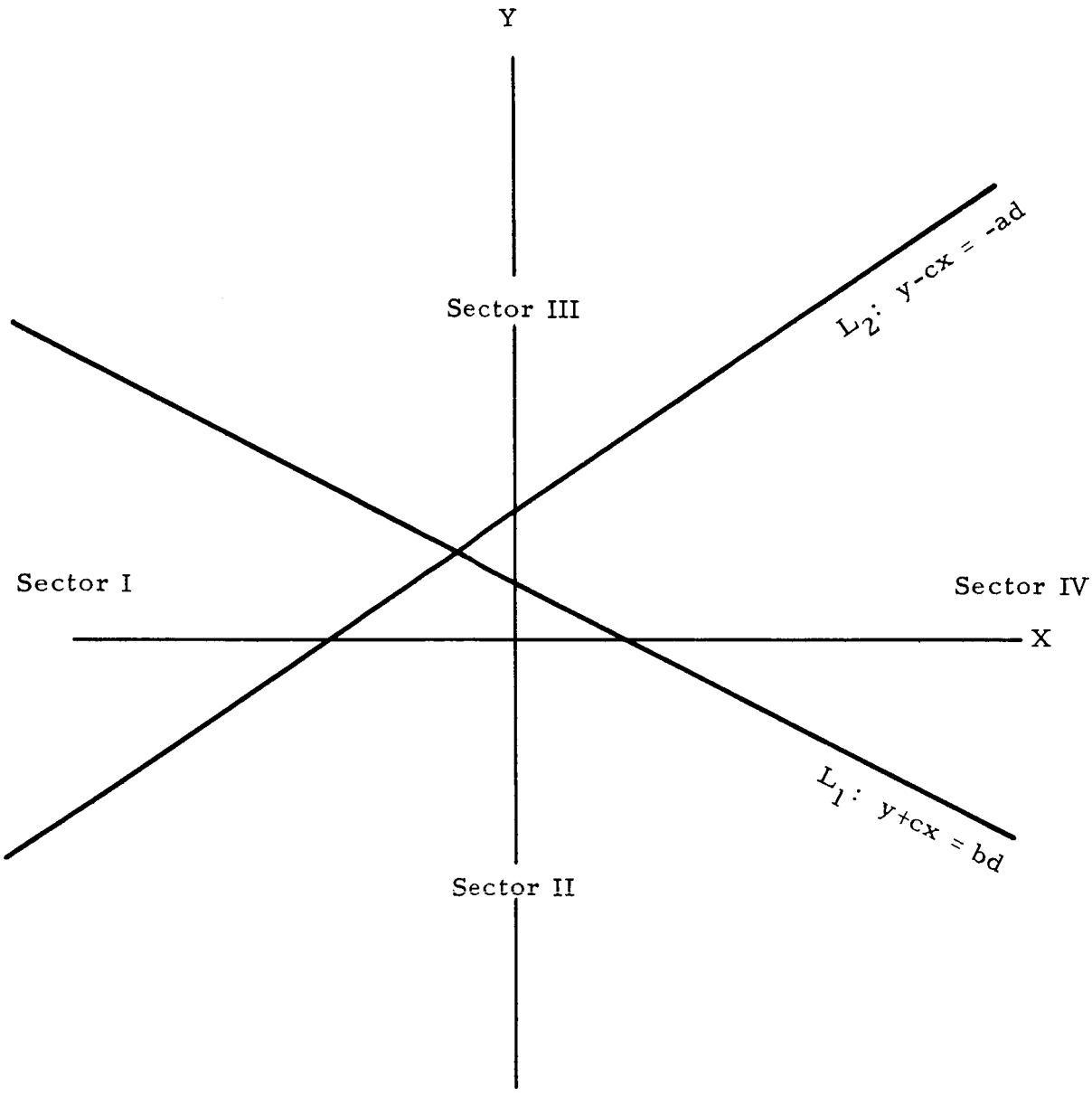


Figure 3-1. Definition of Sectors

$$Q_{jt} = \frac{e^{-\frac{b^2}{2}}}{2\pi} \int_{-\frac{\pi}{2}}^{\tan^{-1}\frac{c^2(b-a) - b - a}{2bc}} e^{-\frac{b^2}{2}\tan^2\Psi} d\Psi$$

$$+ \frac{e^{-\frac{a^2}{2}}}{2\pi} \int_{\tan^{-1}\frac{c^2(b-a) + a + b}{2ac}}^{\frac{\pi}{2}} e^{-\frac{a^2}{2}\tan^2\Psi} d\Psi$$

Substituting $z = \tan\Psi$

$$Q_{jt} = \frac{e^{-\frac{b^2}{2}}}{2\pi} \int_{-\infty}^{-\frac{a-\rho b}{b\sqrt{1-\rho^2}}} \frac{e^{-\frac{b^2}{2}z^2}}{z^2+1} dz$$

$$+ \frac{e^{-\frac{a^2}{2}}}{2\pi} \int_{\frac{b-\rho a}{a\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{a^2}{2}z^2}}{z^2+1} dz$$

and letting $u = bz$

$$Q_{jt} = \frac{be^{-\frac{b^2}{2}}}{2\pi} \int_{\frac{a-\rho b}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{u^2+b^2} du$$

$$+ \frac{ae^{-\frac{a^2}{2}}}{2\pi} \int_{\frac{b-\rho a}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{u^2+a^2} du \qquad (14)$$

A necessary and sufficient condition for the origin to fall in Sector II is that $b > 0$ and $a < 0$. It can be shown analytically that Q_{jt} is again defined by Equation 14. It is to be noted that the second integral is in a negative direction, that is, $\tan^{-1}c(b-a)/b+a$ is greater than $\tan^{-1}c$ in Equation 13.

By the symmetry of the problem and answer, the case where $a > 0$ and $b < 0$ (which defines Sector III) is readily seen to produce the form given in Equation 14.

The probability for the origin in Sector IV, defined by $a < 0$ and $b < 0$, is obtained by subtraction. The forms already obtained are sufficient to provide the probabilities of Sectors I, II, and III when the origin is in Sector IV. Define

$$F(a,b,\rho) = \frac{be^{-\frac{b^2}{2}}}{2\pi} \int_{\frac{a-\rho b}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{u^2+b^2}\,du$$

so that $Q_{jt} = F(a,b,\rho) + F(b,a,\rho)$ when the origin is in Sectors I, II, or III. The substitution

$$x' = -x$$

$$y' = -y$$

shows that the probability of Sector I is

$$P_I = F(-a,-b,\rho) + F(-b,-a,\rho)$$

when the origin is in Sector IV. (To demonstrate this, turn Figure 3-2 upside down.) Similarly, the substitutions

$$x' = -y$$

$$y' = x$$

and

$$x' = y$$

$$y' = -x$$

give

$$P_{II} = F(-a,b,-\rho) + F(b,-a,-\rho)$$

and

$$P_{III} = F(a,-b,-\rho) + F(-b,a,-\rho)$$

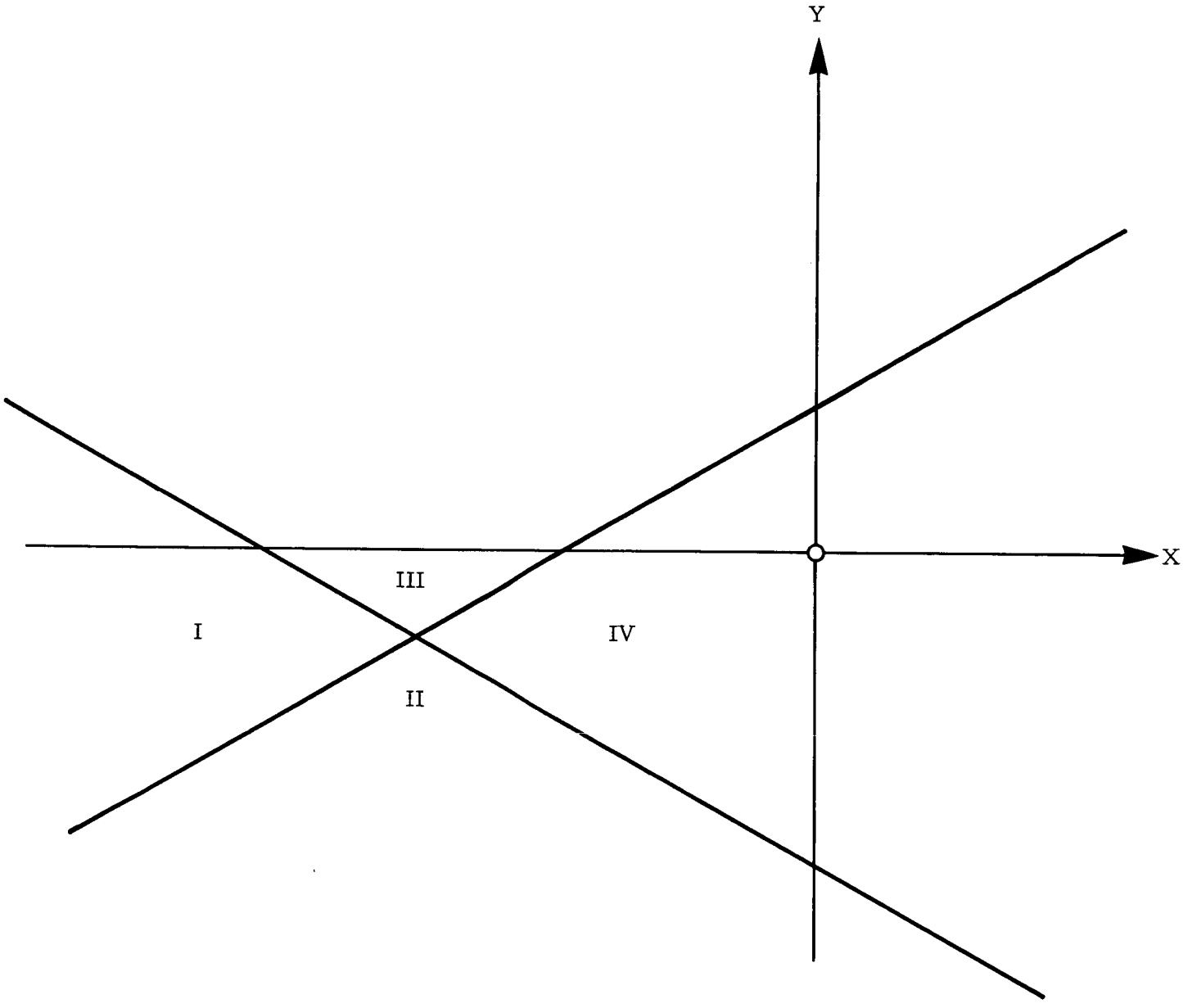


Figure 3-2. Configuration for Origin in Sector IV

respectively. Noting that

$$F(a, -b, -\rho) = -F(a, b, \rho)$$

we have for the origin in Sector IV

$$Q_{jt} = 1 - P_I - P_{II} - P_{III}$$

$$= 1 - F(-a, -b, \rho) - F(-b, -a, \rho) - F(-a, b, -\rho) - F(b, -a, -\rho)$$

$$- F(a, -b, -\rho) - F(-b, a, -\rho)$$

$$= 1 - F(-a, -b, \rho) - F(-b, -a, \rho) + F(-a, -b, \rho) + F(b, a, \rho)$$

$$+ F(a, b, \rho) + F(-b, -a, \rho)$$

$$Q_{jt} = 1 + F(b, a, \rho) + F(a, b, \rho)$$

Thus the form of Q_{jt} for Sector IV is distinguished only by the addition of 1.

There are five boundary cases. These are defined by

(i) $a = 0, b > 0$

(ii) $a > 0, b = 0$

(iii) $a < 0, b = 0$

(iv) $a = 0, b < 0$

(v) $a = 0, b = 0$

Case i $b > 0, a = 0$

In this case, the origin lies on the boundary between Sector I and Sector II.

$$Q_{jt} = \frac{1}{2\pi} \int_{-\tan^{-1} c}^{\tan^{-1} \frac{c(b-a)}{b+a}} \int_{\frac{bd}{c \cos \varphi + \sin \varphi}}^{\infty} e^{-\frac{\gamma^2}{2}} \gamma d\gamma d\varphi$$

$$Q_{jt} = F(a, b, \rho) = F(0, b, \rho)$$

Note that $F(b, 0, \rho) = 0$ when $b \neq 0$; this is a special case of Equation 14.

Case ii $a > 0, b = 0$

By symmetry

$$Q_{jt} = F(0, a, \rho)$$

Case iii $a < 0, b = 0$

In this case, the origin lies on the boundary between Sectors II and IV. In the first integral, the substitution $u = bz$ is invalid. The value of this integral is readily seen to be 1/2, by inspection of Figure 3-1. The result is then

$$Q_{jt} = \frac{1}{2} + F(0, a, \rho)$$

Case iv $b < 0, a = 0$

By symmetry

$$Q_{jt} = \frac{1}{2} + F(0, b, \rho)$$

Case v $a = 0, b = 0$

From Figure 3-1,

$$Q_{jt} = \frac{1}{2\pi} \int_{-\tan^{-1} c}^{\tan^{-1} c} \int_{0}^{\infty} e^{-\frac{\gamma^2}{2}} \gamma d\gamma d\varphi$$

$$= \frac{1}{2\pi} \int_{-\tan^{-1} c}^{\tan^{-1} c} d\varphi$$

$$= \frac{1}{\pi} \tan^{-1} c$$

$$Q_{jt} = \frac{1}{\pi} \tan^{-1} \sqrt{\frac{1+\rho}{1-\rho}}$$

Summarizing:

a	b	Q_{jt}
0	0	$\frac{1}{\pi} \tan^{-1} \sqrt{\frac{1+\rho}{1-\rho}}$
> 0	0	$F(0, a, \rho)$
< 0	0	$\frac{1}{2} + F(0, a, \rho)$
0	> 0	$F(0, b, \rho)$
0	< 0	$\frac{1}{2} + F(0, b, \rho)$
< 0	< 0	$1 + F(a, b, \rho) + F(b, a, \rho)$
All Others		$F(a, b, \rho) + F(b, a, \rho)$

where

$$F(\xi, \eta, \rho) = \frac{\eta e^{-\frac{\eta^2}{2}}}{2\pi} \int_{\frac{\xi - \eta\rho}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{u^2 + \eta^2}\, du$$

3.2 Alternative Algorithms

In this section, several alternative self-organizing algorithms are described. These are compared with the perceptron forced learning rule analyzed in Section 3.1.

3.2.1 Natural Selection

"Natural selection" is a modification of the forced learning rule which produces a decrease of nearly an order of magnitude in the number of logic units required to approach asymptotic performance (see Equation 4 and discussion of Section 3.1). In common with the forced learning rule, it may be considered an open-loop process - that is, the self-organizing routine changes the system parameters in a manner which is not dependent on the system output.

To achieve the hardware efficiency possible with natural selection, the self-organizing routine is simulated on a computer. The forced learning rule is used to design a perceptron which is larger than actually needed to achieve acceptable performance. The forced learning rule assigns large positive weights to logic units which are good indicators that a positive pattern is present, and large negative weights to units which are reliable indicators of negative patterns. Units which are less reliable indicators of the pattern classification are assigned weights of lesser magnitude. The natural selection rule uses these values not only as a measure of the size of the votes given to the logic units in the decision process, but also as measures of the efficiency of these units. Only the most efficient units are included in the hardware realization. In this way, the average efficiency of the units may be increased.

Natural selection has been used in the construction of the Astropower Decision Filter. (Note: the term "decision filter" has been associated with networks designed using natural selection.) In this device, a perceptron of 20,000 logic units was designed in the simulation. Four-hundred logic units were included in the decision filter. The selection of 2% resulted

in an average unit effectiveness in the decision filter of 7-1/2 times the perceptron average. This estimate is based on the closeness of the approach to asymptotic performance. Other examples of the effectiveness of natural selection are given in the next section.

Being an open-loop process, natural selection has little effect on asymptotic performance. Its influence is on the rate at which this performance is approached. It is thus suitable only for those problems amenable to the forced learning perceptron approach.

3.2.2 Iterative Design

"Iterative design" may be considered a closed-loop self-organization process in that the system changes are directly related to the system performance. Such a feedback system is required when the asymptotic performance of a forced learning perceptron is poor due to the high inherent variance (see Equation 4 Section 3.1).

The first successful closed-loop self-organization process was Rosenblatt's error correction rule.* As with forced learning, the process is nonselective and logic unit efficiency is low. Error correction is, however, quite successful in coping with problems unsuitable for forced learning, but at the price of a considerably more complex self-organizing routine.

Iterative design is a selective closed-loop process, including in the network only such units as substantially contribute to system performance. As such, it produces much more efficient networks than the error correction process. It must be accomplished in a computer simulation. It is described here in conjunction with a technique for generating logic units using discriminant analysis. Together, discriminant analysis and iterative design provide an extraordinarily powerful self-organizing algorithm, as will be shown by some empirical results. Since this is the most effective algorithm available today, it is described in considerable detail.

The first-level logic units map the signal space into a "recognition" space. The second-level logic units, or decision elements, partition the recognition space with hyperplanes. These simple boundaries in the recognition space have a more complex polyhedral representation in the signal space.

*Rosenblatt, F., Principles of Neurodynamics, Sparton Books, 1962.

A hardware limitation places basic constraints on the choice of the linear logic unit structure. Circuit design state of the art limits the accuracy and stability of placement of the partitioning hyperplanes. This requires that logic units in the first level be limited to subspaces of low dimensionality in which relatively high accuracy of hyperplane placement may be maintained. It further requires that the recognition space (typically large in number of dimensions) defined by the output of the first-level logic units widely separate the pattern classes so that minor inaccuracies in the placement of the hyperplanes will not be a major source of error.

In order to maximize the separation of pattern classes in the recognition space, the design approach under discussion selects the first-layer logic units sequentially. Each unit added to the machine is selected from a population of possible logic units. The criterion for selection is the improvement over existing interclass separations.

It is necessary to specify a population of alternative subspaces, since computational difficulties prevent the determination of the optimum subspace to be partitioned by the linear logic unit. However, once given a subspace, discriminant analysis can be used to determine a useful hyperplane within it.

Thus the present technique involves (1) random selection of a set of subspaces of the sample space (random to achieve maximum originality), (2) determination of hyperplanes in these subspaces by discriminant analysis, and (3) selection of a logic unit on the basis of the improvement in the separation of pattern classes that it effects. It should be noted that the design iteration includes steps (2) and (3), since the addition of a logic unit modifies the interclass separations, and hence the most suitable subspace hyperplanes.

The selection procedure requires that the separability of the pattern classes be quantized. Furthermore, the quantization should be intimately related to the restrictions of linear partitions. In the technique under discussion, this is accomplished by means of a loss function.

The technique requires the existence of a representative set of classified patterns. Let the hyperplane H be defined by the equation

$$\sum_j w_j x_j - \theta = 0$$

where x_j is the output of the j-th logic unit. Each pattern is assigned a loss, $L(P, w, \theta)$. The loss numbers are functions of the locations of the patterns with respect to the hyperplane and the classification of the patterns. The loss associated with the hyperplane H is the sum of these loss numbers over all patterns in the representative set. The separability of the pattern classes is described in terms of the minimum value of this loss over all hyperplanes.

$$\text{Separability}^{*} = f\left[\inf_{w,\theta} \sum_i L(P_i, w, \theta)\right]$$

The current design technique selects for inclusion in the network that logic unit which provides the greatest increase in the separability - that is, which provides the greatest decrease in the minimum loss.

Many forms of the loss function are possible - polynomial, linear, integral, exponential, and combinations. In the examples of Tables 3-4 through 3-7 (given later in this section) an exponential form of the loss function was utilized. As before

$$\delta^i = \begin{cases} 1 & \text{If the i-th pattern is in the positive class} \\ -1 & \text{If the i-th pattern is in the negative class} \end{cases}$$

Then the loss function used had the form

$$L(P_i, w, \theta) = e^{-\delta_i (\sum_j w_j x_j(i) - \theta)}$$

The loss thus depends exponentially on the distance of the point representing P_i in the recognition space from the separating hyperplane (for a given length of the defining weight vector w).

This particular form of the loss function provides an exponential rate of change of the loss number as the distance of the pattern from the hyperplane changes. This ensures that the greatest emphasis will be

---

*For some forms of the loss function $L(P_i, w, \theta)$ (for example, the exponential form), the existence of a hyperplane which separates the sample patterns perfectly implies (1) separability = f(0), (2) the corresponding hyperplane is not unique, and (3) there is no vector w_o such that

$$\sum_i L(P_i, w_o, \theta) = \inf_{w,\theta} \sum_i L(P_i, w, \theta).$$

A normalization on the length of w is suggested.

placed on the most poorly separated patterns. The function has little to recommend it computationally, however, as it makes exact determination of the weights impractical.

In the examples, an approximation of the optimizing vector was utilized. The approximation is accomplished in two parts: (1) a relatively accurate approximation to the optimum weight vector is determined for the space defined by the logic units already selected, and (2) each space created by the addition of a prospective logic unit to this space has a hyperplane defined by simple modification of the original hyperplane.

The augmentation of the space introduces a new dimension, x_{n+1}, and the association of a weight and threshold to this dimension produces the augmented hyperplane

$$\sum_{j=1}^{n} w_j x_j + w_{n+1} x_{n+1} = \theta + \Delta\theta_{n+1}$$

The values of w_{n+1} and $\Delta\theta_{n+1}$ are chosen to optimize the separability function

$$f = -\Sigma \left[\left(e^{-\delta_i \left(\sum_{j=1}^{n} w_j x_j(i) - \theta \right)} \right) \left(e^{-\delta_i \left(w_{n+1} x_{n+1}(i) - \Delta\theta_{n+1} \right)} \right) \right]$$

This optimization may be readily accomplished.

To produce the higher order approximation after the selection of a logic unit, the same basic procedure is applied iteratively to the existing coordinates of the recognition space, modifying the hyperplane by optimizing with respect to the weight for each coordinate. In the examples given, this cycle was repeated four times whenever a new coordinate was added to the recognition space.

Discriminant analysis* is a multivariate statistical technique applicable to certain taxonomic problems. For classification problems which meet the underlying assumptions, the technique provides an optimum decision rule for assigning observed vectors, or patterns, to one of several classes.

*See, for example, C. R. Rao, Advanced Statistical Methods in Biometric Research, John Wiley & Sons, Inc., New York, 1952.

The assumptions basic to discriminant analysis are as follows:

a. The observation P is a vector chance variable. It is drawn from one of several multivariate normal distributions, depending only on its class membership.

b. These distributions differ only in their (unknown) mean vectors. An (unknown) covariance matrix is common to all of the distributions.

c. The decision criteria are to be based on sample vectors drawn randomly from each of the distributions.

A major result of the analysis is a set of linear discriminant functions for the pattern vector, one for each pattern class. The values assigned to a pattern by these functions are called L-scores. The decision rule assigns the pattern to the pattern class corresponding to the highest L-score. For the specific case of two pattern classes, the decision criterion reduces to a single separating hyperplane.

For the design problem under discussion, the classifications required of the decision elements are dichotomous, even when more than two pattern classes are involved. Therefore, this technique is suggested for the establishment of hyperplanes within the randomly selected property filter subspaces.

The assumptions underlying appropriate use of the discriminant analysis technique generally are too severe to be met in most pattern recognition problems. Notwithstanding, there are several reasons why its use is indicated. First, the technique is quite robust, and provides good results in many cases where the assumptions are only approximately met. Secondly, it seems unlikely that the hyperplanes generated by this technique would not, in general, be of greater utility than randomly placed hyperplanes. And finally, the selection process provides a means for removing anomalous cases.

In determining the property filter hyperplanes for the examples presented in Tables 3-4 through 3-7, the sample patterns were weighted. Each pattern was given a weight proportional to the current loss

assigned to it by the loss function. Each time a logic unit was selected for inclusion in the network, the weights were revised and new property filter hyperplanes were generated. The population of logic units was thus kept responsive to the varying requirements of the growing decision network.

The validity of the iterative design technique has been tested by comparison of designs which it achieves with forced learning perceptrons and natural selection decision filters designed from the same unit populations for the same pattern recognition problems.

A set of 80 hand-printed letters was generated (consisting of 20 S's, H's, A's, and N's), and the activity of each of 50 randomly chosen logic units for each pattern was determined. The logic units were chosen to have four excitatory and two inhibitory inputs, each selected with uniform probability from a 35-element rectangular input space. The activities were determined under the rule that the unit was considered active if the number of active excitatory inputs exceeded the number of inhibitory inputs by two or more. An input was considered active for a pattern if any part of the pattern fell on the square corresponding to that input.

For the perceptrons, the decision filters, and the networks designed utilizing only the basic iterative process and a single-stage approximation of the optimal hyperplane, the computations were performed manually. The small problem size was necessitated by the amount of manual processing required, but was deemed sufficient to allow comparison of the design techniques. The extended iteratave technique, involving the use of discriminant analysis to specify logic units and a two-stage iterative process to approximate the decision hyperplane, required an amount of calculating large enough that a digital computer was considered necessary even for the small problem.

The simplest problem chosen was the discrimination of S's from H's. Table 3-4 indicates that the iterative design technique produced a better design than the other alternatives, even when alternative networks contained more logic units.

Table 3-5 shows the comparative efficiency of the experimental networks for three problems of differentiating pairs of letters. Table 3-6 gives the error rates of designs for these three problems achieved by the discriminant analysis iterative technique after the selection of each additional

logic unit to the network. The results of network design to separate the four classes of patterns through the use of two classification units are shown in Table 3-7.

3.2.3 Iterative Design for Sequential Computers

The design techniques discussed up to this point are specifically suited to parallel logic devices. When conditions are such that it is expedient to use a digital computer for pattern recognition in difficult problems (ones requiring many logic units), the design algorithms should be tailored to the special features of sequential operation.

The primary simplification which arises may be understood by reference to the iterative design technique. It will be remembered that operation of a logic unit in this mode is considered to divide the ordered set of sample patterns into two subsets: one corresponding to points on the positive side of the hyperplane defined by the logic unit, and one to points on the negative side. The iterative technique is primarily a method of recombining these subsets so that the ordering of the resultant set is more easily separable into predetermined classes.

In sequential operation it is neither necessary nor normally desirable to recombine these subsets. When operation of one subprogram of a pattern recognition program has placed a pattern as occurring in a particular region of the pattern space, the next subprogram should be one suited to discrimination in this region. Thus operation of a sequential pattern recognition program deals primarily with the selections of finely tailored subprograms on the basis of increasingly accurate estimates of the location of patterns under investigation. The process terminates when the location has been specified finely enough that all patterns in the defined cell are homogeneous in classification.

From the aspect of the logic designer, the design of any subprogram reduces to the specification of a partition on the input space which maximally separates the subset of patterns which are expected to refer to that subprogram. Thus, as shown in Figure 3-3, if the first subprogram divides the space into two regions, then subprogram 2 partitions the same space to separate the set of positive classified patterns, while subprogram 3 is based only on the negatively classified patterns. The outputs of subprogram 2 result

TABLE 3-4

DISCRIMINATION OF S's AND H's

Type of Network	No. of Errors
50-Unit Perceptron	1
10-Unit Decision Filter (Difference Weights)	1
5-Unit Decision Filter (Difference Weights)	2
10-Unit Decision Filter (Logarithmic Weights)	1
5-Unit Decision Filter (Logarithmic Weights)	2
5-Unit, Iteratively Designed Network	None

TABLE 3-5

DISCRIMINATION OF PAIRS OF PATTERNS

Type of Network	No. of Errors			
	S&H vs A&N	S&A vs H&N	S&N vs H&A	Total
50-Unit Perceptron	3	5	12	20
8-Unit Decision Filter (Difference Weights)	5	8	10	23
8-Unit Decision Filter (Logarithmic Weights)	5	7	6	18
8-Unit, Iteratively Designed Network	2	5	6	13
1-Unit Machine (Discriminant Analysis Property Filter)	4	5	10	19

TABLE 3-6

DISCRIMINATION OF PAIRS OF PATTERNS
ITERATIVE DESIGN USING DISCRIMINANT ANALYSIS

No. of Units in Network	No. of Errors		
	S&H vs A&N	S&A vs H&N	S&N vs H&A
1	4	5	10
2	4	2	6
3	1	0	2
4	0	0	0

TABLE 3-7

DISCRIMINATION OF S vs H vs A vs N

Type of Network	No. of Errors		
	Classification Unit 1 S&H vs A&N	Classification Unit 2 S&N vs H&A	Network
50-Unit Perceptron	3	12	14
8-Unit Decision Filter (Difference Weights)	4	13	15
8-Unit Decision Filter (Logarithmic Weights)	5	11	13
8-Unit, Iteratively Designed Network	3	9	11
8-Unit Discriminate Analysis Iterative Design	0	0	0

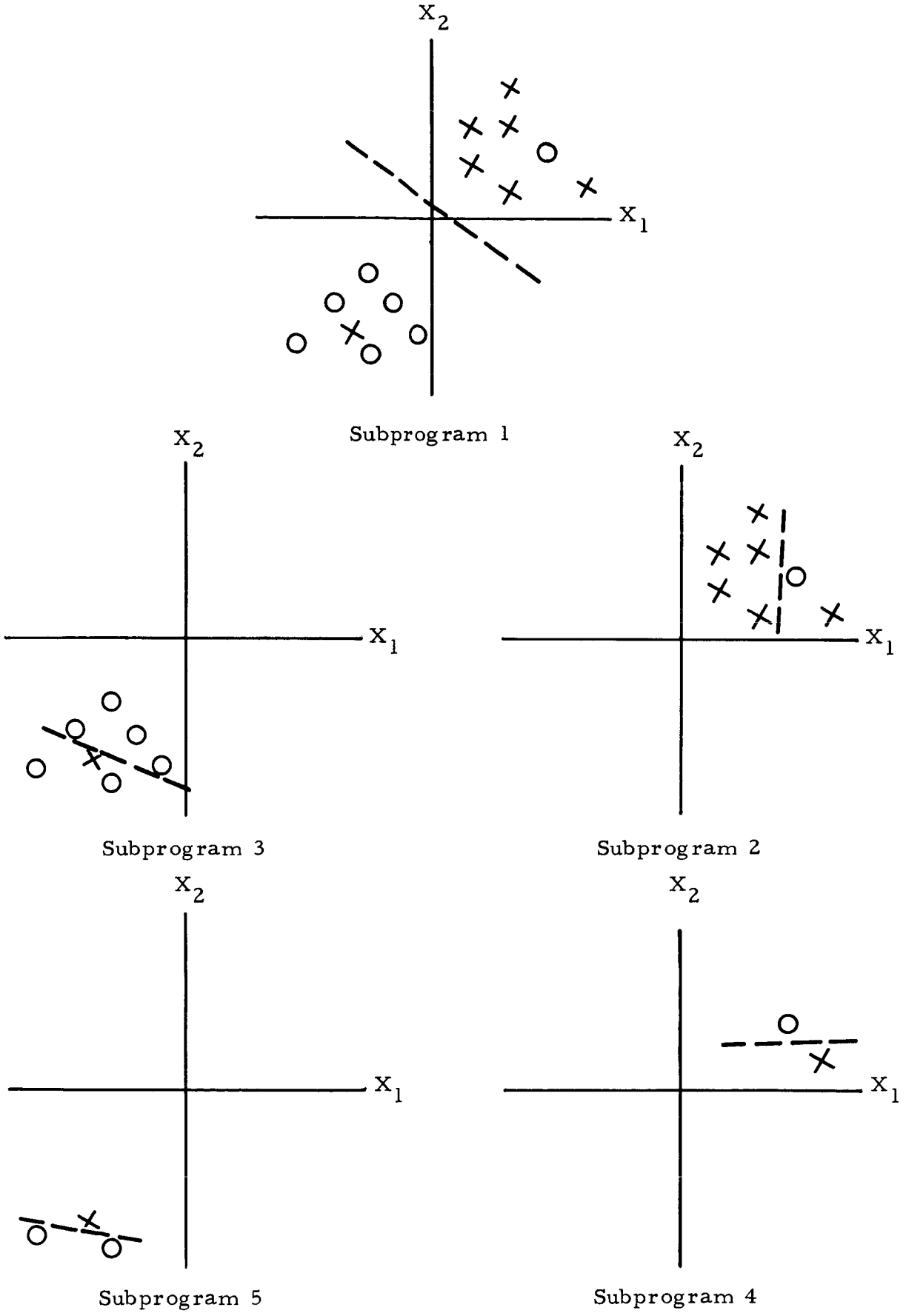


Figure 3-3. Operations Typical of Consecutive Subprograms in Sequential Pattern Recognition

either in the identification of the pattern as positive or in the selection of subprogram 4, which resolves the remaining uncertainty. Similarly, subprogram 3 either classifies the pattern or calls further subprograms designed to reduce the remaining uncertainty.

Succeeding reductions in the size and diversity of the sets of patterns to be classified implies that succeeding subprograms will tend to be more accurate. The basic problem of how and what kind of partition to design for each subprogram remains. The solution at this point is a confession of ignorance. Difficult problems are those in which the shapes of pattern class regions are too complex to lend themselves to linear separation. The only more complex separations with clearly defined design algorithms are polygonal discriminants (particularly useful are the polygonal discriminant surfaces defined by the iterative design technique). Therefore, the suggested design approach for subprogram logic is a variation of the iterative design technique.

A variation on the basic technique is suggested primarily because iterative design is based primarily on a binary response unit. This approach is not optimal for computer-based operation because it restricts the number of subprogram referrals available as outputs, and in so doing discards much of the information obtained by the subprogram. This is particularly important in the early parts of the problem, where relatively small numbers of first-layer decisions will be unable to effect much unscrambling of patterns of different classifications. A single threshold in such cases will produce comparatively little information on pattern classifications, while multiple thresholds may very greatly reduce problem complexity. It should be noted that the precision available in digital computer calculations reduces the variance expected in location of thresholds and patterns with respect to thresholds, thereby reducing the primary basis for objection to multiple thresholding.

Multiple threholding of a single response unit is clearly not a final answer. While considerations of information theory can probably be applied to such a scheme to determine criteria for useful placement of these thresholds, the criteria for design of this response vector is not optimum multiple separations. One immediately appealing approach would be the use of iterative design techniques for the selection of a small transformation space of high separability, followed by use of cluster analysis and canonical reduction

techniques to achieve very high separation in multiple output logic. The utility of such approaches, of course, must be verified by experimental evidence before any real import is attached to them.

To restate the foregoing arguments, proper use of a digital computer for pattern recognition requires that its sequential nature be taken into consideration. Therefore the system under consideration bases its choice of operations on the information it has obtained, discarding as little as possible. This can be expected to require the design of more logic than would be necessary for a parallel machine, but to require fewer operations per final decision (adding inexpensive tape memory to reduce expensive machine time).

The individual subprograms to be designed should have multiple outputs, which are used to identify patterns as belonging to particular subclasses. The subprograms do this by partitioning the input space with polygonal surfaces, which are particularly suited for discrimination of the subclass of patterns which select the subprogram.

While this decision scheme is recognized as not being optimal in any sense, it is offered as an easily defined mechanism, superior to linear discriminants for difficult problems, and almost certainly superior to direct simulation of parallel logic for computer operation.

4.0 OPTICAL CORRELATION MEASUREMENTS

4.1 Introduction

In this section the optical correlation measurements are discussed in detail. Paragraph 4.2 describes the procurement and selection of the photographic raw material. Paragraph 4.3 discusses the preparation of the slides suited for insertion in the optical correlator. The logbook design which specified the sequence of the measurements, is discussed in 4.4. The design, testing, and calibration of the optical correlator, as well as the measurement routine, are described in 4.5.

4.2 Procurement of Duplicates of TIROS 35mm Film

The video signal transmitted from TIROS satellites during their radio contact with the ground station is displayed on a cathode ray tube. A 35mm camera takes photographic negatives of each individual scope display. The film rolls are sent to the National Weather Record Center (NWRC) in Ashville, N. C., for storage and production of second generation copies for interested parties.

Astropower, working on an in-house study program, had obtained 12 rolls of TIROS I photographs from NWRC previous to submission of the Cloud Pattern Interpretation proposal to NASA Headquarters. A close inspection of all these photographs revealed that not enough distinct vortex patterns were available on these 12 rolls. In anticipation of the contract award and wary of delays, Astropower had ordered 20 rolls of TIROS III film as early as September 5, 1962. Up to May 23, 1963, only two passes (each roll contains a large series of passes) had been received from NWRC. However, with the cooperation of Mr. H. Lowell of NASA Headquarters and Mr. H. Oseroff of Goddard Space Flight Center, more than 1200 photos from TIROS V and TIROS VI were obtained. Not only did these negatives show a greater percentage of storm formations, but in addition they were of superior quality, being first-generation copies sent to Astropower directly from the TIROS station manager on Wallops Island. These TIROS V and TIROS VI photos have been considered exclusively for this task. A second bulk of film rolls, received from Wallops Island in July 1963, could not be incorporated because the program was too far advanced.

The preselection was achieved as follows. The weather maps contained in the TIROS catalogs* were screened for signs of vortices. These signs had been inserted in the maps by the meteorologist who interpreted the cloud photographs. The map also contained information on the pass number during which the pictures associated with the map had been taken. The numbers of the passes encountering vortices were recorded and ordered. Only a part of the order was received by Astropower. Of the film rolls received all photographs taken during these passes were inspected. All photos that appeared to show the familiar vortex pattern were selected. For each photo showing a vortex, a second was selected showing no obvious storm formation. A total of 260 photos was obtained in this fashion. These were taken to the consulting meteorologist, Professor M. Neiburger of U. C. L. A., for inspection, and all photos that could not clearly be classified as vortex or nonvortex were isolated.

A last selection was made considering the area of earth coverage within the photo. It is conceivable that the selected vortex photos could show, in general, more horizon in a certain picture area than would nonvortex photos, or vice-versa. This would lead to an incorrect interpretation of the classification clues. Since the intent was to discriminate between storm and no-storm, it was decided to exclude horizon areas by selecting only a round disc within each rectangular negative. This resulted in the additional benefit that one picture of a pair could be rotated with respect to the other without changing the overlap area, thus allowing additional correlation measurements with little added effort. Table 4-1 gives a list of the negatives in the vortex class. The first column signifies the numbers given to the particular negatives for easy identification during the optical correlation work. Since the same vortex pattern is often seen from different angles, there often appears a sequence of photographs in the same row. The second column contains the readout pass number; the third column

* 1. Catalog of TIROS V Cloud Photography for June 1962.
2. Catalog of TIROS V Cloud Photography for July 1962.
3. Catalog of TIROS V Cloud Photography for Aug 1962.
4. Catalog of TIROS V-VI Cloud Photography for Sept 1962.
All produced by the National Weather Satellite Center, Washington, D. C.

contains the pass number during which the vidicon camera took the picture. A "D" in the next column signifies direct transmission, i.e., the cloud scene was transmitted during the observation, while "T" signifies that the picture was stored on tape before transmission. The following columns contain the date of observation and the geographical location of the vortex pattern. All pictures were taken with the TIROS V or TIROS VI wide angle camera system and were received, displayed, and photographed at the TIROS Command and Data Acquisition Station on Wallops Island.

4.3 Preparation of Slides

It was necessary that the selected 35mm cloud cover negatives be converted to slides usable on the optical correlator. The rotatable slide holders contained in the correlator setup were designed to hold a disc of 55mm outside diameter. To obtain the necessary mechanical rigidity, the slide frames were fabricated from 13-gauge hard aluminum (QQA661T6). An inner diameter of 1 in. was chosen for the transparency area. The frames were manufactured to $\pm$2/1000-in. tolerances. After punching they were tumble-burred and anodized. Two marks were put on the frames; a hairline serving as a reference mark for insertion into the optical holder, and a semicircular notch to define the rotation in a jig that was used during the photographic processing. A total of 400 frames was procured.

As described in Section 3.1.4, an identical pair of slides for each picture was required to find its mean gray level and its gray level variance. To obtain identical pairs, unexposed photographic material* was bonded to the frames. After overnight drying, the frames were inserted one at a time in a jig mounted on the enlarger base. Then three frames were exposed in sequence to the enlarged image** of the original 35mm negative.*** Since exposure time, brightness, development time, and

---

* Eastman Kodak Commercial Film with Estar Base

** An enlargement factor of 1.54 was chosen. In this way no horizon appears on the slides.

*** The photographic processing was accomplished under a subcontract with Dean Hesketh Photographical Laboratory in Anaheim, Calif.

TABLE 4-1

LIST OF VORTEX NEGATIVES SELECTED FOR OPTICAL CORRELATION MEASUREMENTS

Number	Pass of Readout	Pass of Picture Taking	Direct or Tape	TIROS No.	Date	Location of Vortex
1 through 6	1053	1053	T	V	31 Aug '62	19.8 N 134.6 E
7	1151	1151	D	V	7 Sept '62	either 54.5 N 73.5W* or 38.5 N 57.5 W
8	684	683	T	V	6 Aug '62	39 S 135 E
9 through 12	670	669	T	V	5 Aug '62	34.5 S 130.5 E
13 through 15	1147	1146	T	V	7 Sept '62	38.5 N 57.5 W
16 through 20	1133	1132	T	V	6 Sept '62	61.5 N 17.5 E
21 through 24	1010	1010	T	V	28 Aug '62	44 N 160 E
25 and 26	1204	1204	D	V	11 Sept '62	48 N 79.5 W
27 through 31	598	597	T	V	31 July '62	52 S 142 E
32 and 33	1137	1137	D	V	6 Sept '62	52 N 77 W
34 through 36	807	807	D	V	14 Aug '62	44.5 N 62.5 W
37 and 38	954	953	T	V	25 Aug '62	either 20.5 N 131 E* or 29.5 N 136 E
39 through 43	145	144	T	VI	28 Sept '62	45 N 32 E
44 through 49	921	920	T	V	22 Aug '62	56 N 34 W
50	598	597	T	V	31 July '62	52 S 142 E

* During these passes two vortices were observed. Since the location of the vortex pattern is only of secondary interest for our work, no attempt was made to clarify this ambiguity.

developer liquid were held as nearly constant as possible, it was hoped to obtain at least one identical pair out of every triplet. Unfortunately, the variations in the photographic material and those due to tolerances in the mentioned parameters were large enough to cause variations in the measurable transparency of the slides. It was not considered economically feasible to attempt to achieve higher precision in the photographic processing technique. Therefore, the "identical pair" criterion was established realistically by specifying that the two slides of a pair should not differ by more than 5% in their measured transparency. Seven hundred slides were produced in four sets with three to seven slides per frame to achieve the required 100 identical pairs. These 100 pairs then formed the imagery for the optical correlation measurements.

4.4 Logbook Design for Optical Measurements

With 50 slide pairs of the vortex class and 50 slide pairs of the nonvortex class yielded by the photographic processing, there existed a combination of 5000 picture pairs that could be correlated against each other. Since the rotation of a slide in the image plane results in another pattern with respect to the envisioned recognition system, the number of possible pattern combinations could be increased indefinitely. Estimating the time required to accomplish a single measurement on the optical correlator and considering the time and money allotted to the program, it was decided that a total of approximately 10,000 measurements would be made. These measurements would provide a reasonable estimate of the overlap areas between the pattern classes as required for the mathematical analysis of Section 3.0.

The following types and numbers of measurements were required:

a. Transparency of each individual slide. This measurement was taken on a total of 700 slides. Only 200 of these slides were selected for the succeeding measurements.

b. Transparency of an exactly aligned identical pair. The 100 pairs produced 100 measurements of this type.

After these measurements were performed, only one slide of each pair was used for further measurements.

c. It was decided that each negative should be correlated to 16 other negatives of its own class and 16 of the opposite class, thus resulting in 1600 pairs. Each pair was to be measured for six different rotations relative to each other, yielding 9600 measurements. The rotation of a slide was readily achieved on the optical setup, yielding a convenient method for obtaining additional data points with little effort.

A total of 10,400 measurements was therefore required. A logbook was designed which enumerated the individual picture pairs to be correlated. The vortex negatives were numbered 1 to 50, the nonvortex negatives 51 to 100. A 100 x 100 square matrix was then plotted (see Figures 4-1 and 4-2), and individual squares were marked. A mark at the crossing of column m with row n indicates that negatives m and n form one of the picture pairs to be correlated. The design was started with the upper left quarter (established by the overlap of the first 50 rows with the first 50 columns), which signifies vortex-vortex correlations. After excluding the diagonal from 1-1 to 50-50, marks were inserted at random, starting at the upper left corner, with the restrictions that each column and each row should contain 16 marks, and that the resulting pattern should be symmetrical about the diagonal. The same procedure was applied to the lower two quarters using the pattern already established in the first quarter and reshuffling numbers at random.

Portion B (Figure 4-1) consists of nonvortex-nonvortex correlations, while portion C indicates vortex-nonvortex correlations. (Only portion A is shown in detail in Figure 4-2). The resulting design assured a good spread of picture combinations.

From the detailed charts A, B, and C, the actual logbook was compiled for the 9600 measurements discussed under (c) above. The logbook has the following format:

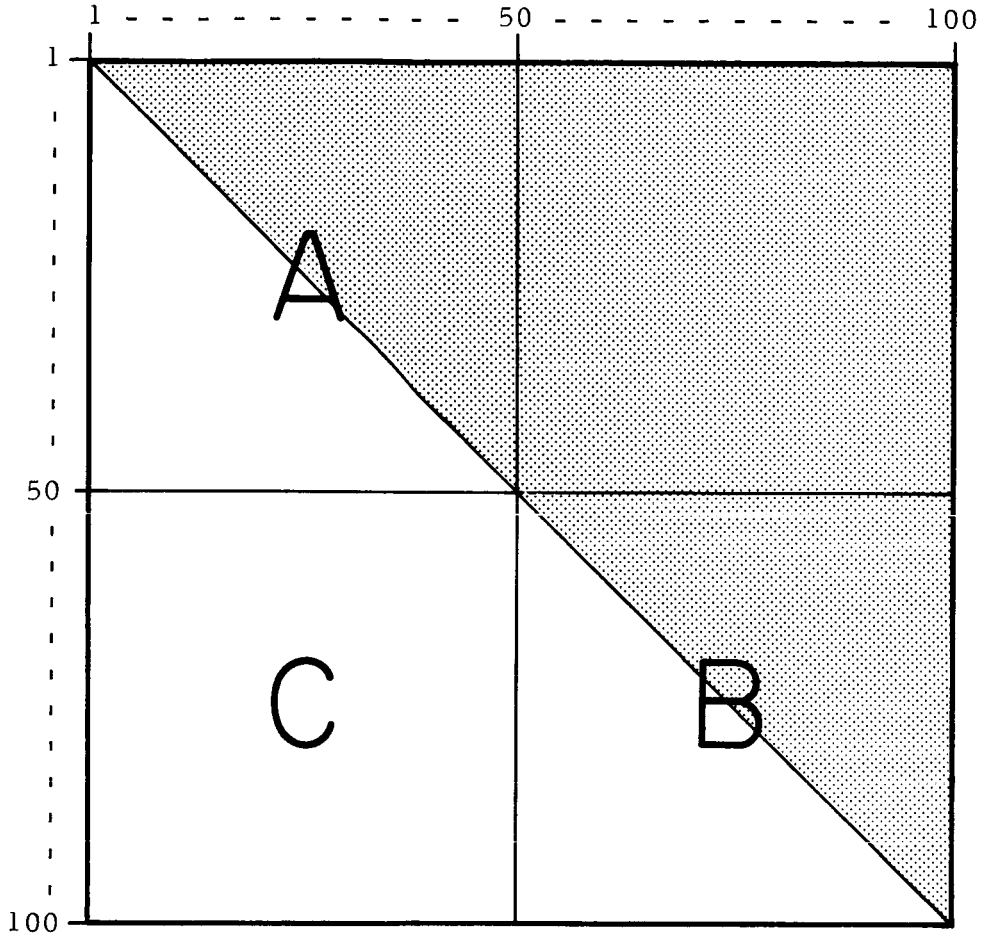


Figure 4-1. Organization of the 100 x 100 Matrix Used for the Logbook Design (The shaded area is the mirror image of the clear area.)

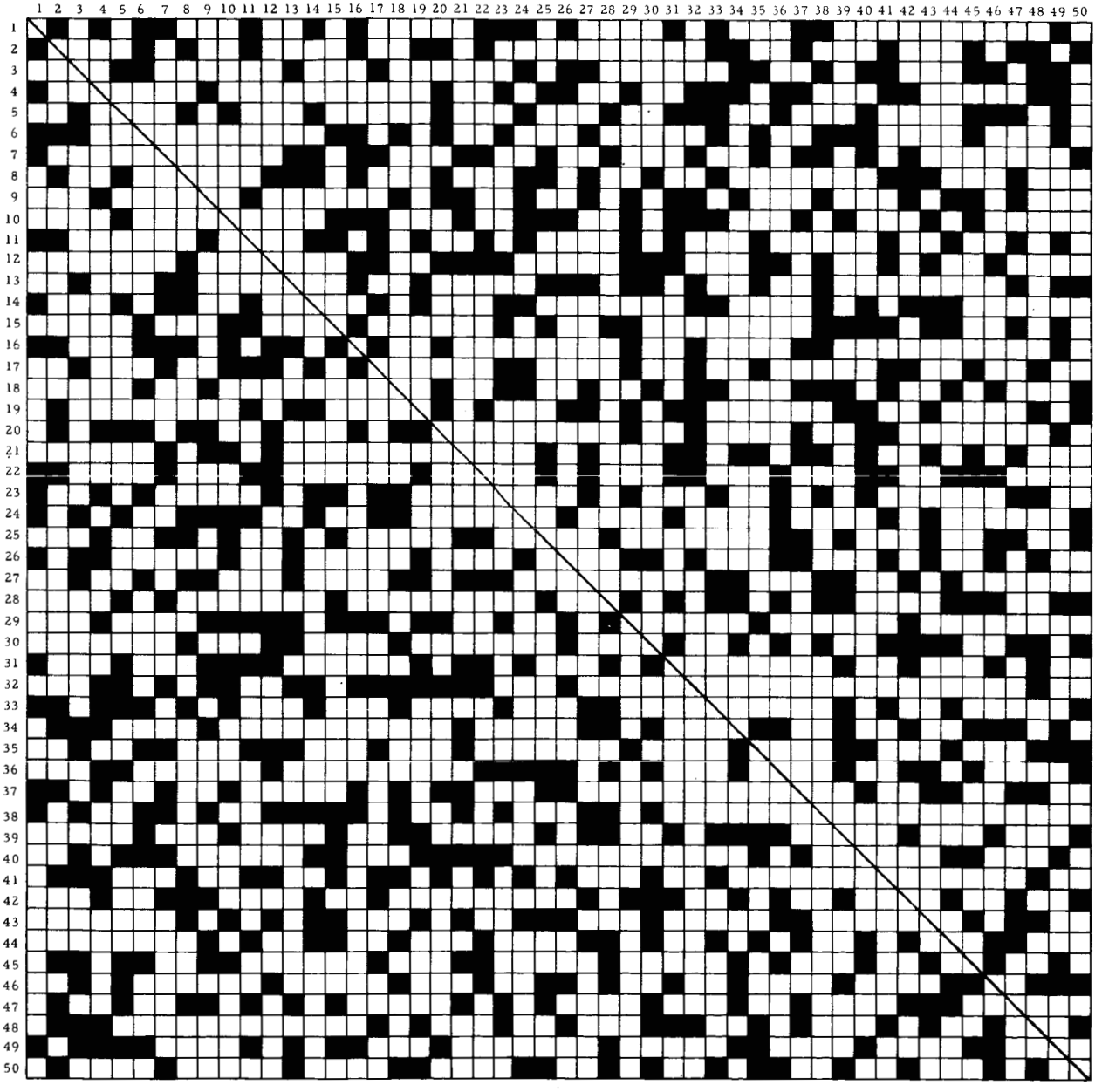


Figure 4-2. Detailed Layout of the A-Portion of the Square Matrix Shown in Figure 4-1.

Measurement No.	Slide X	Slide Y	Combined Transparency for a Mutual Rotation of					
			0^{o}	60^{o}	120^{o}	180^{o}	240^{o}	300^{o}

This layout was chosen so the data could be transferred directly onto IBM cards by the key punch operator.

4.5 Optical Correlator and Optical Measurements

4.5.1 Description of Equipment

4.5.1.1 Functional Description

A system schematic of the optical correlator is shown in Figure 4-3. Figures 4-4 and 4-5 give an overall view of the equipment as used in making the slide measurements. The modulated light beam is generated in the small black enclosure mounted on the left end of the optical bench. The diverging beam emerges from an aperture in the front of the source enclosure and passes through a collimating lens. The collimated beam then passes through the transpariencies mounted in slide holders 1 and 2 respectively. After passing through the slide holders, the collimated beam is decollimated and focused upon the photo detector. The output of the photo detector is monitored by an oscilloscope and measured by a tunable microvoltmeter. The reading of the microvoltmeter is directly proportional to the transmittance of slides inserted into the holders.

4.5.1.2 Modulated Light Source

A view of the light source with cover removed is shown in Figure 4-6. The 6v, wound filament, prefocused lamp is energized by a high stability, regulated direct current power supply.

The light from the lamp follows two paths. A mirror is positioned over the lamp to reflect light through openings in the two front panels into an externally mounted photo detector housing. The beam is positioned so that the chopper wheel modulates the beam before it strikes the lead sulfide photo detector. The output of the photo detector is

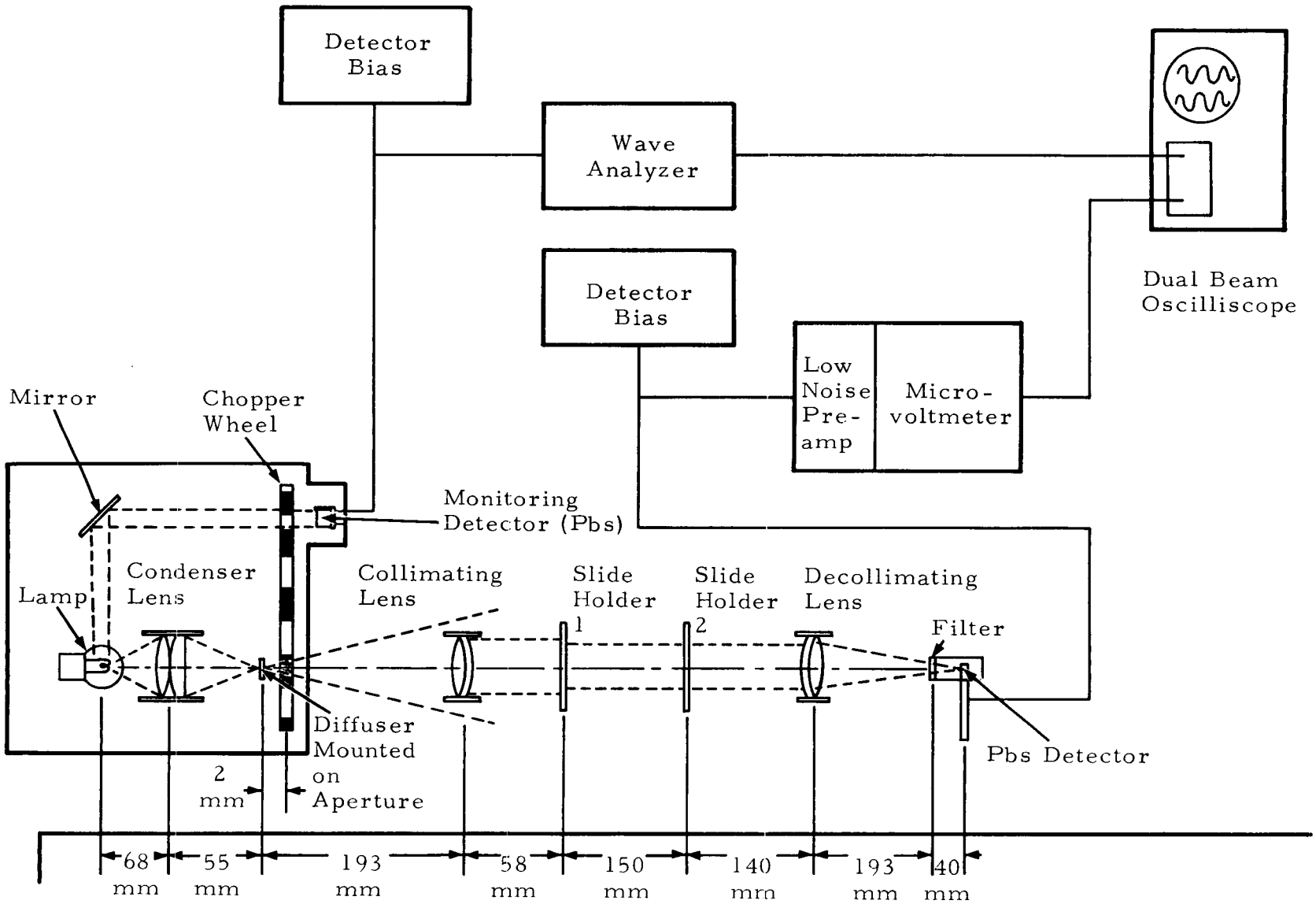


Figure 4-3. System Schematic

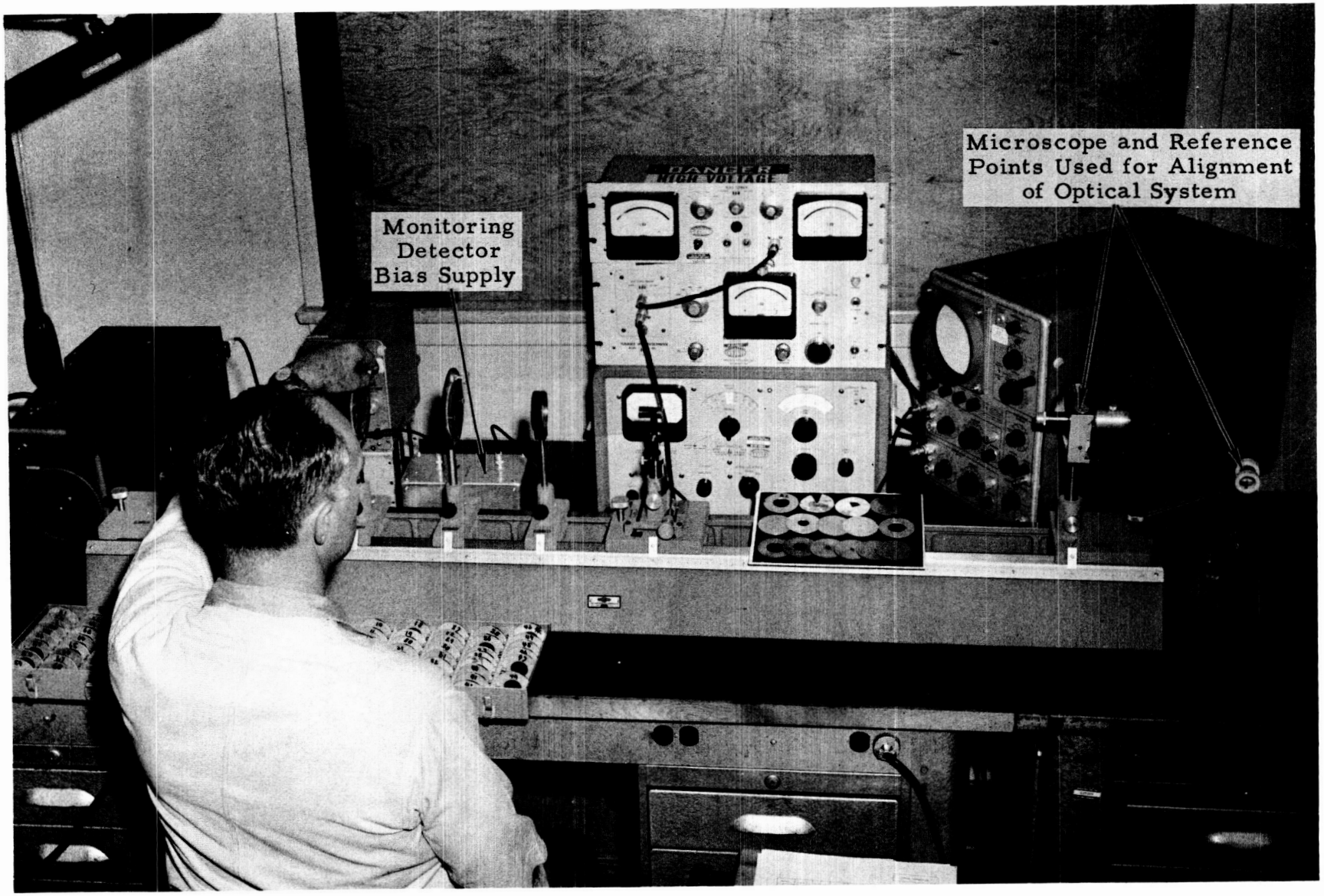


Figure 4-4. Overall View of System in Operation

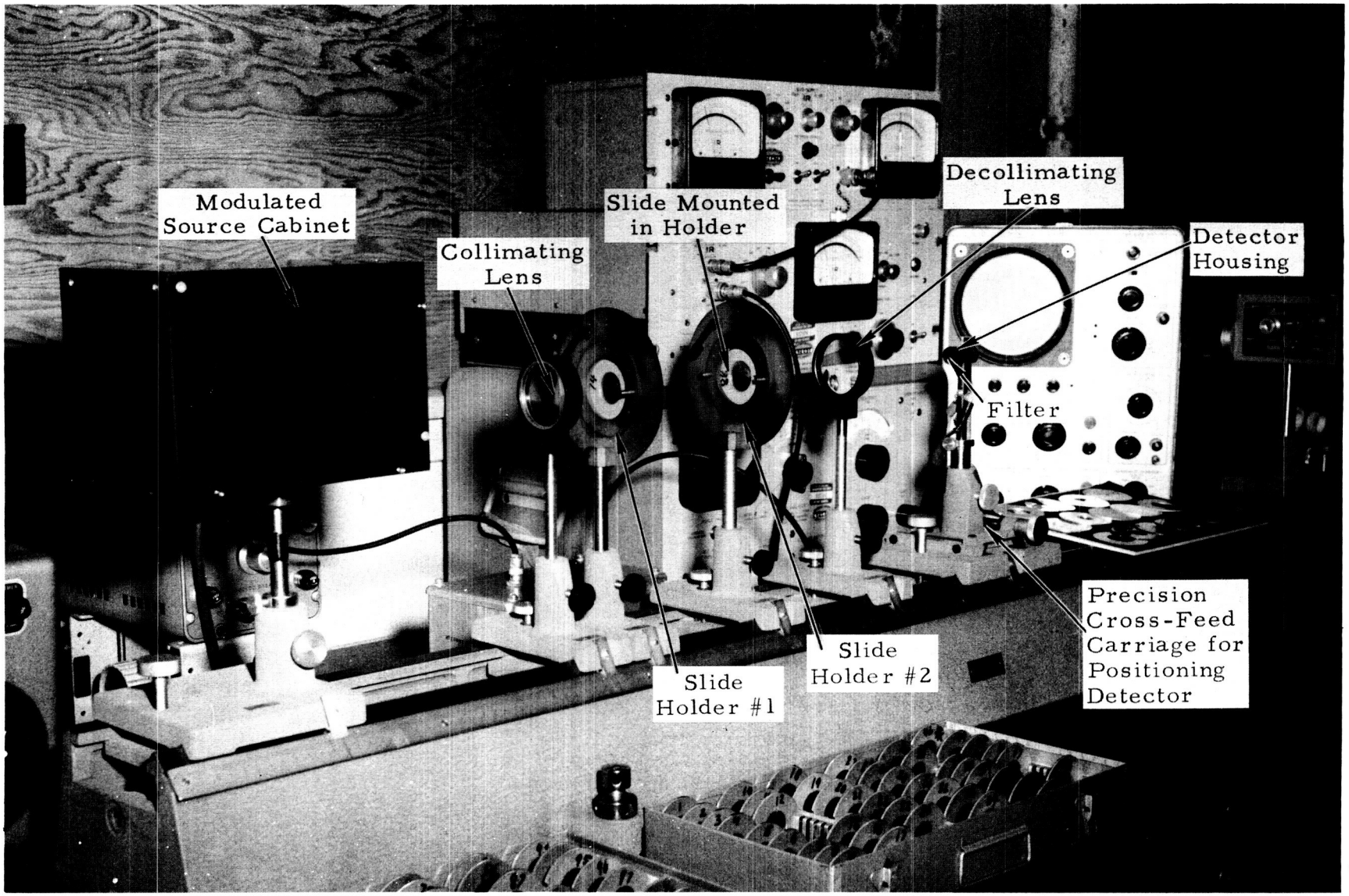


Figure 4-5. Arrangement of Components on Optical Bench

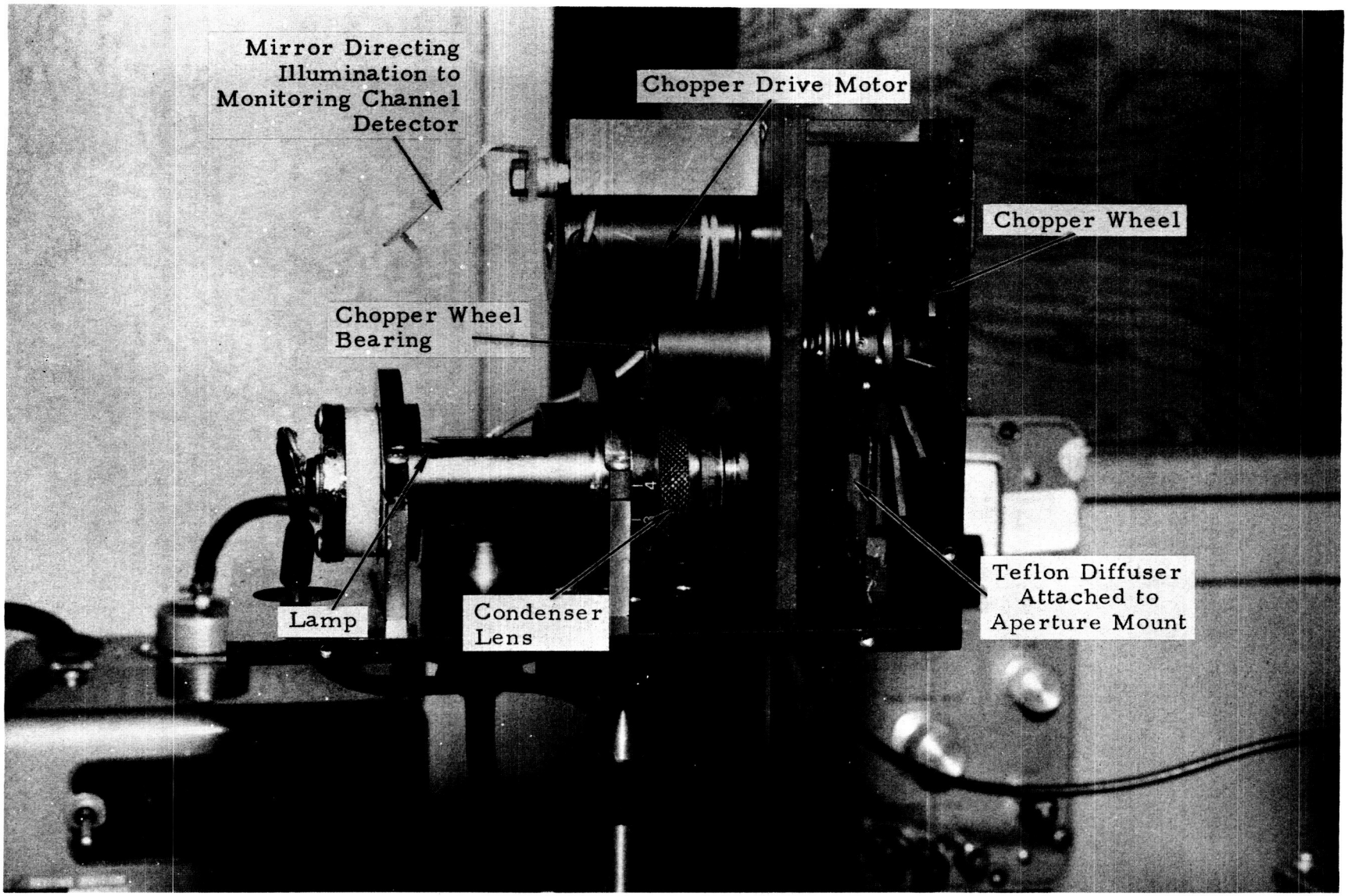


Figure 4-6. Modulated Radiation Source

measured by a tunable microvoltmeter and monitored by an oscilloscope. Thus the lamp intensity and the chopper wheel speed are continuously monitored. The chopper wheel is driven by a hysteresis synchronous motor through a belt and pulley arrangement, and provides a very stable modulating frequency after the stabilization of a small, preliminary, thermally-induced drift. The chopper wheel configuration is shown in Figure 4-7.

The primary beam from the lamp passes through a condensing lens, then a sheet of diffusing material, and finally a 0.1 mm aperture in the image plane of the condenser. Two millimeters from the aperture is the rotating chopper wheel, which modulates the beam after it passes through the aperture.

4.5.1.3 Optical Bench and Components

The modulated beam passes through the exit aperture in the front of the source enclosure and illuminates the collimating lens. A sketch of the arrangement of the optical components is shown in Figure 4-3, and a description of the lenses may be found in the equipment list (4.5.1.6). All optical components are centered on the optical axis defined by the center of the source aperture and parallel to the ways of the optical bench. The film slides are mounted so that the centers of the film portions that are to be measured remain on the optical axis if the slide is rotated. The slides are supplied with an index mark on the aluminum disc so they can be positioned accurately in the holders. Figure 4-8 shows two of the slides mounted in the holders so that the light beam passes through both slides. The hand of the operator is seen rotating the stage of the first slide holder.

The collimated beam, after passing through the films in the two slide holders, is focused on the detector by the decollimating lens.

4.5.1.4 Detector

The beam first passes through a filter before illuminating the detector. The filter has two functions: to reduce stray light and spurious radiation that would affect the sensitivity of the detector,

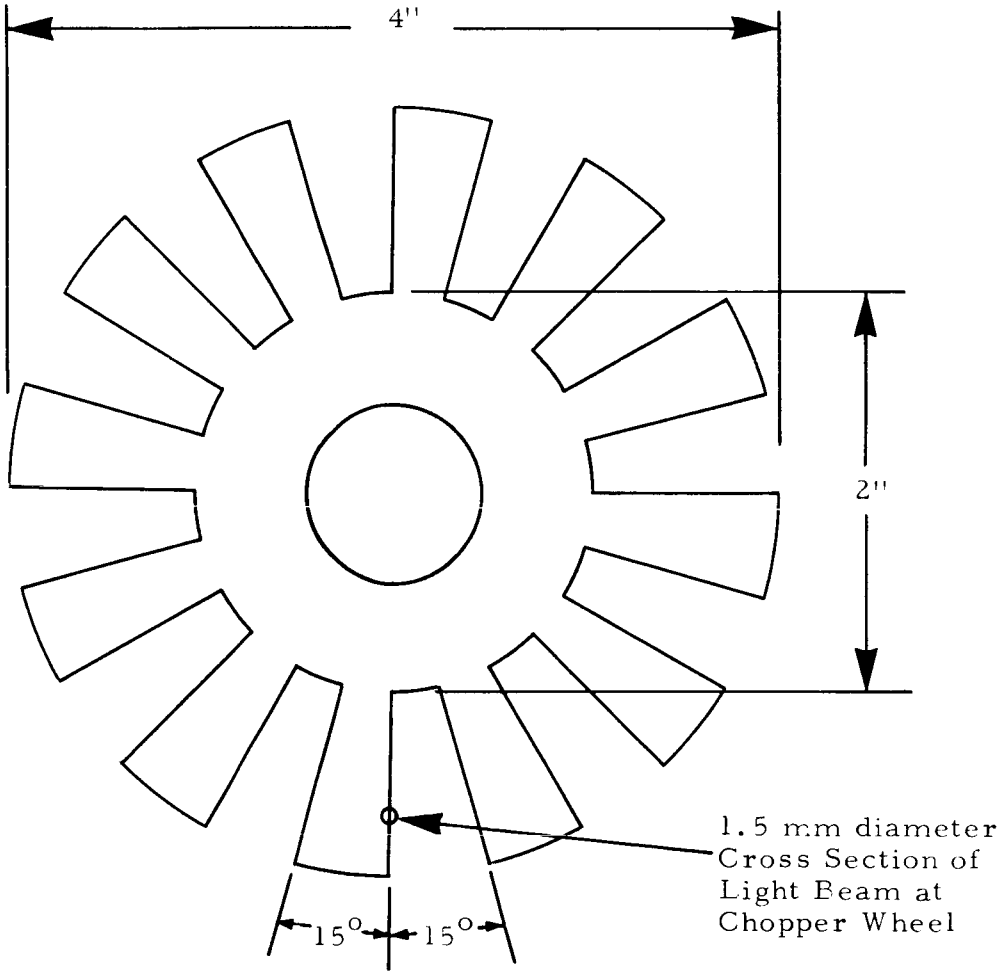


Figure 4-7. Chopper Wheel Configuration

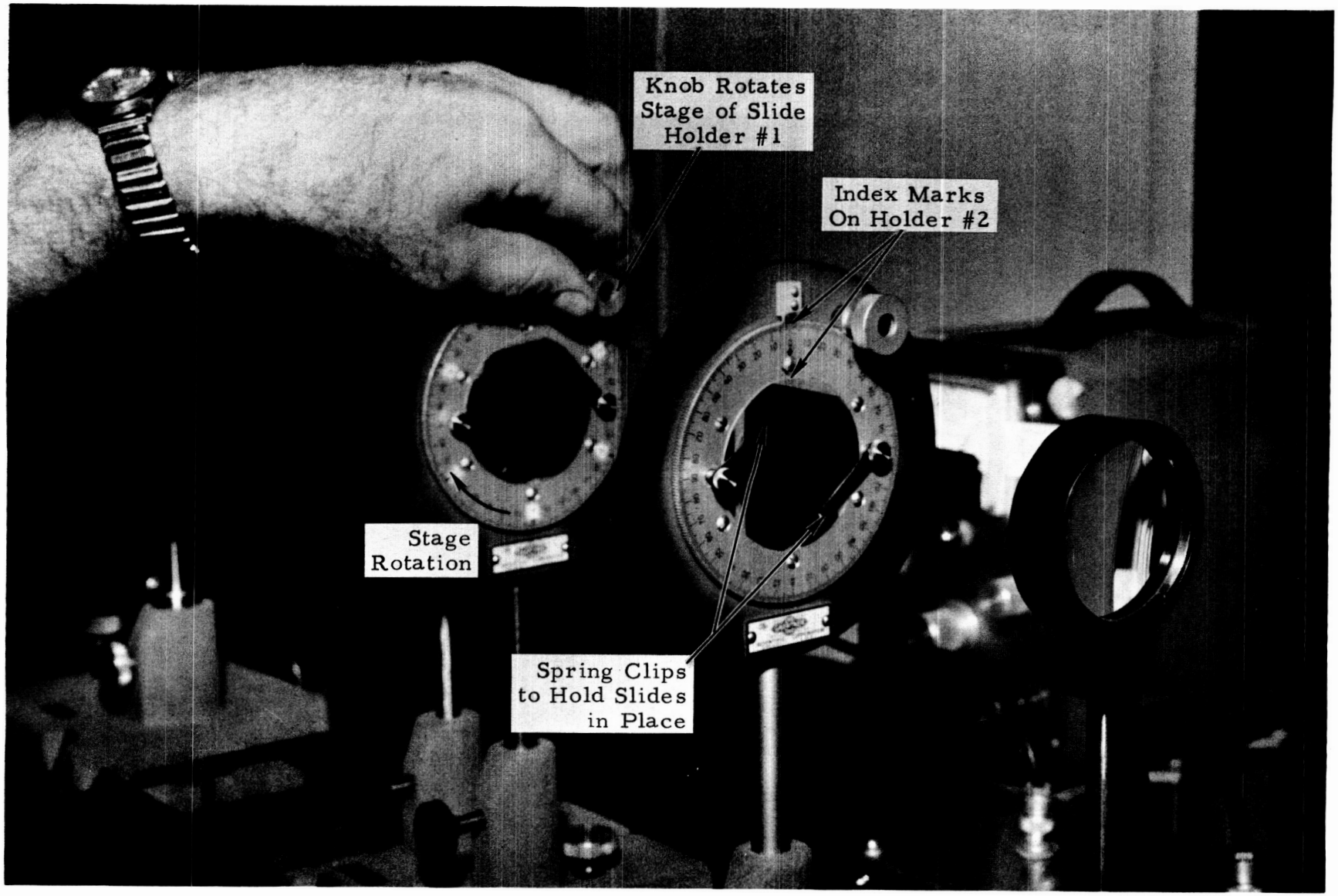


Figure 4-8. Operator Rotating Slide in Holder No. 1

and to limit the radiation falling on the detector to the band which was chosen for optimum operating characteristics of the system. The detector is a lead sulfide photoconductor with a 0.06 x 0.06 in. sensitive area. The sensitive area has been covered with a 0.05 x 0.05 x 0.002 in. Teflon chip for use as a diffuser. The detector is attached by polystyrene cement to a 5/8 in. polystyrene rod and is provided with a slip-on hood containing the filter. A sketch of the arrangement is shown in Figure 4-9. The detector is not cooled, but operates in a controlled environment of 69°F ±3°F and relative humidity of 55% ±8%. The 5v detector bias is supplied by a high impedance adjustable bias supply with provisions for continuous monitoring of the bias voltage and current.

4.5.1.5 Amplifier and Meter

The detector output is connected to a tunable microvoltmeter with a low noise preamplifier input. The microvoltmeter is adjusted to the chopper frequency of 240 cps and a bandwidth of 9.6 cps. The voltage measured by the microvoltmeter is displayed on the upper channel of a dual beam oscilloscope; the trace of the monitoring channel is displayed on the lower beam so that the correct frequency relationship is readily ascertained. The indicated meter range of the microvoltmeter is 1 μv to 10,000 μv but may be readily expanded by adjusting the fine gain control and using a scaling factor determined by calibrating the new readings. By this technique an effective linear range of 150 to 500,000 μv was achieved.

4.5.1.6 Monitoring Channel

The detector used in the monitoring channel is a lead sulfide photo detector identical to that used in the main data channel. It is mounted externally in a machined aluminum housing on the front of the source housing, as shown in Figures 4-4 and 4-10. Bias is supplied to the detector by a fixed 6v battery supply. The detector output is connected to a wave analyzer, which is basically a tunable microvoltmeter. The wave analyzer was tuned to the 240 cps modulating frequency, and the source output was monitored by both the wave analyzer meter and by an oscilloscope connected to the output of the wave analyzer. The monitoring signal was expanded vertically and displayed so that a change in signal amplitude of 2 parts per 1000 could easily be detected.

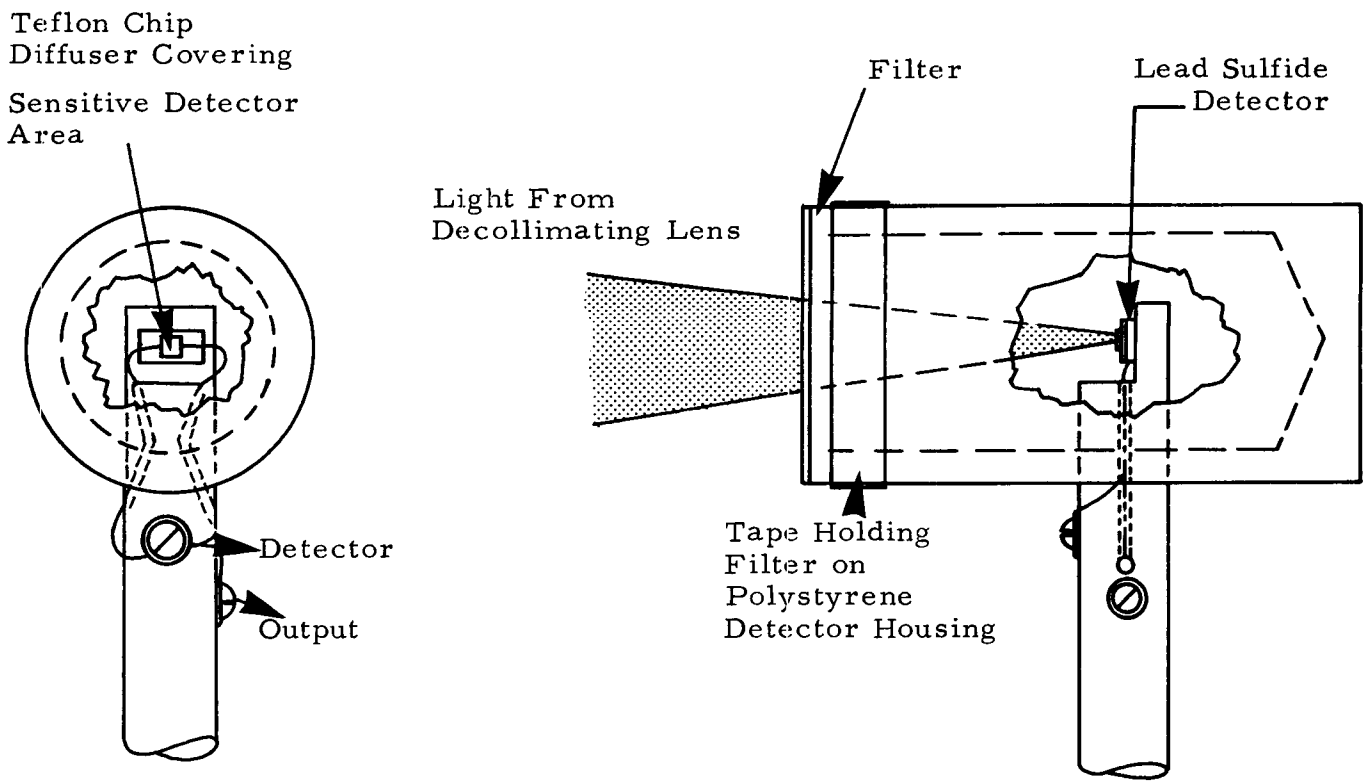


Figure 4-9. Measurement Channel Detector Configuration

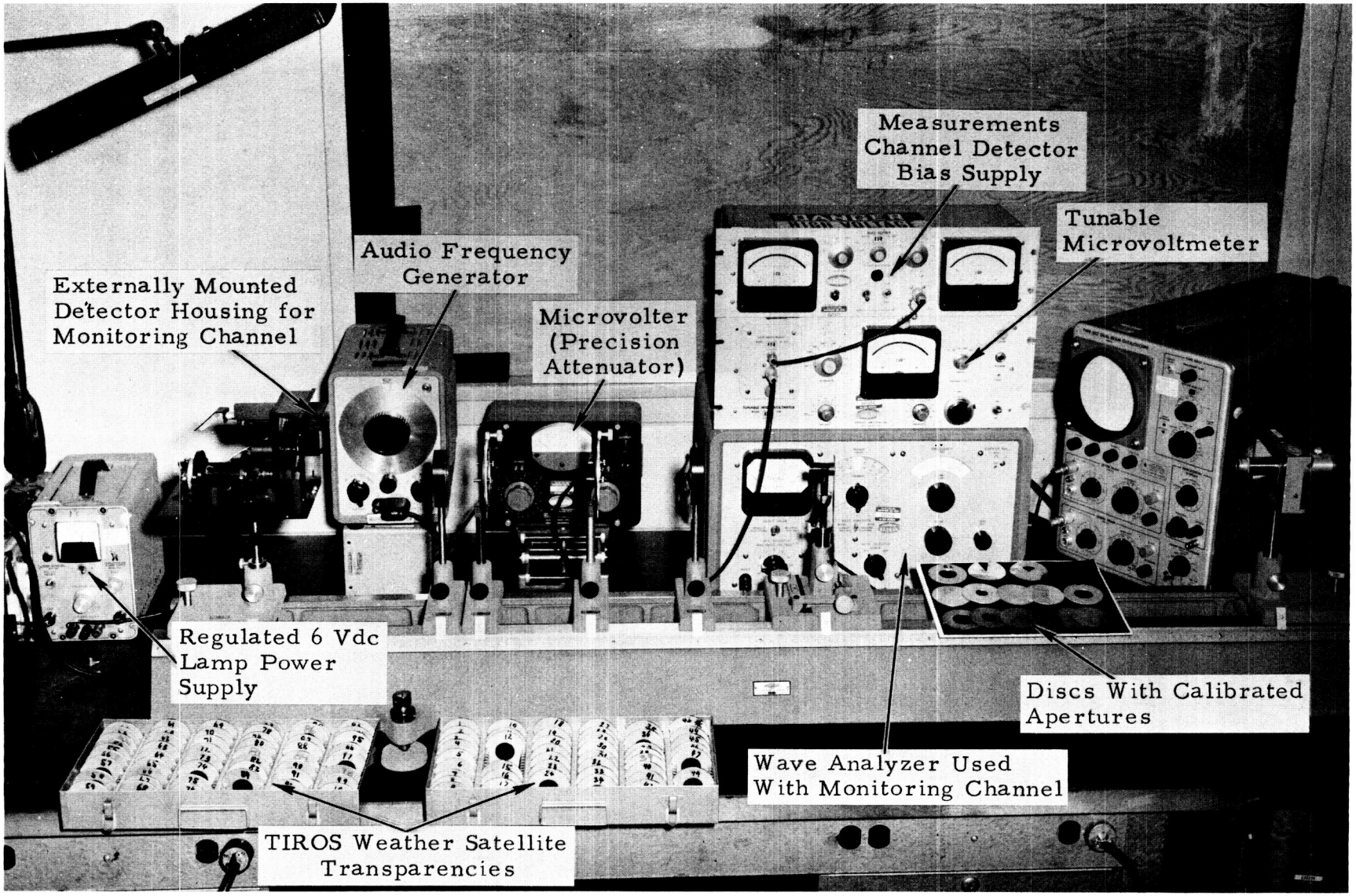


Figure 4-10. Equipment for Calibrating System

4.5.1.7 Equipment List

The following equipment was incorporated in the optical correlator:

Optical Bench - Gaertner 160 cm Model L360N optical measurements system

Detectors - IR Industries photo detector; Lead sulfide, Type SA4; Substrate, 0.250 x 0.150 in.; sensitive area, 0.06 x .06 in.

Optical Components

Lamp - 6v prefocused

Condenser lens - 1 in. diameter, 28 mm focal length; Kellner eyepiece, coated

Collimating lenses - 52 mm diameter, 193 mm focal length, Achromatic Doublet (coated)

Chopper wheel drive motor - Globe Industries 75A 119-2, hysteresis, 115 vac - single-phase, 60 cycle, synchronous speed, 1200 rpm

Electronic Equipment

Measurement channel bias - IR Industries Model ISL 502 bias supply

Microvoltmeter - IR Industries Model 600 tunable microvoltmeter with IR Industries Model ISL 6001 low noise, plug-in preamplifier unit

Monitoring channel microvoltmeter - Hewlett Packard Model 302A wave analyzer oscilloscoep

Oscilloscope - Tektronix Type 502 dual beam

Calibration Equipment

Hewlett Packard Model 200AB audio oscillator

General Radio Co. audio frequency microvolter

4.5.2 Difficulties Encountered in Meeting the System Requirements

It was necessary that the system meet certain requirements to insure obtaining valid data suitable for computer analysis. These requirements were:

a. That the light passing through the two slides be collimated and have uniform flux density

b. That the system be linear over the range of measurements made; i.e., given a particular increase in the light transmitted by the slides, a proportional increase in the meter reading should result

c. That the output of the detector be constant for all rotational positions of any single slide, all other parameters to be held constant

d. That the output of the source be constant both in intensity level and in modulation frequency

All of the above requirements were met after considerable developmental work. The greatest difficulty was encountered in establishing a beam of constant flux density. After considerable experimentation with heated nichrome strips, various filament configurations, and several diffuser materials, a uniform flux density was obtained. A small sheet of 2-mil textured vinyl plastic was mounted at the aperture in the source housing. Figure 4-11 shows the relative flux densities of the collimated beam as finally achieved. The values are presented as measured at the start and at the completion of the optical correlation program. Figure 4-12 shows the flux density locations with respect to the optical bench setup.

4.5.3 Calibration Procedures

4.5.3.1 Calibration of Microvoltmeter

Figure 4-10 shows the equipment used for calibrating the microvoltmeter. The audio frequency generator was tuned

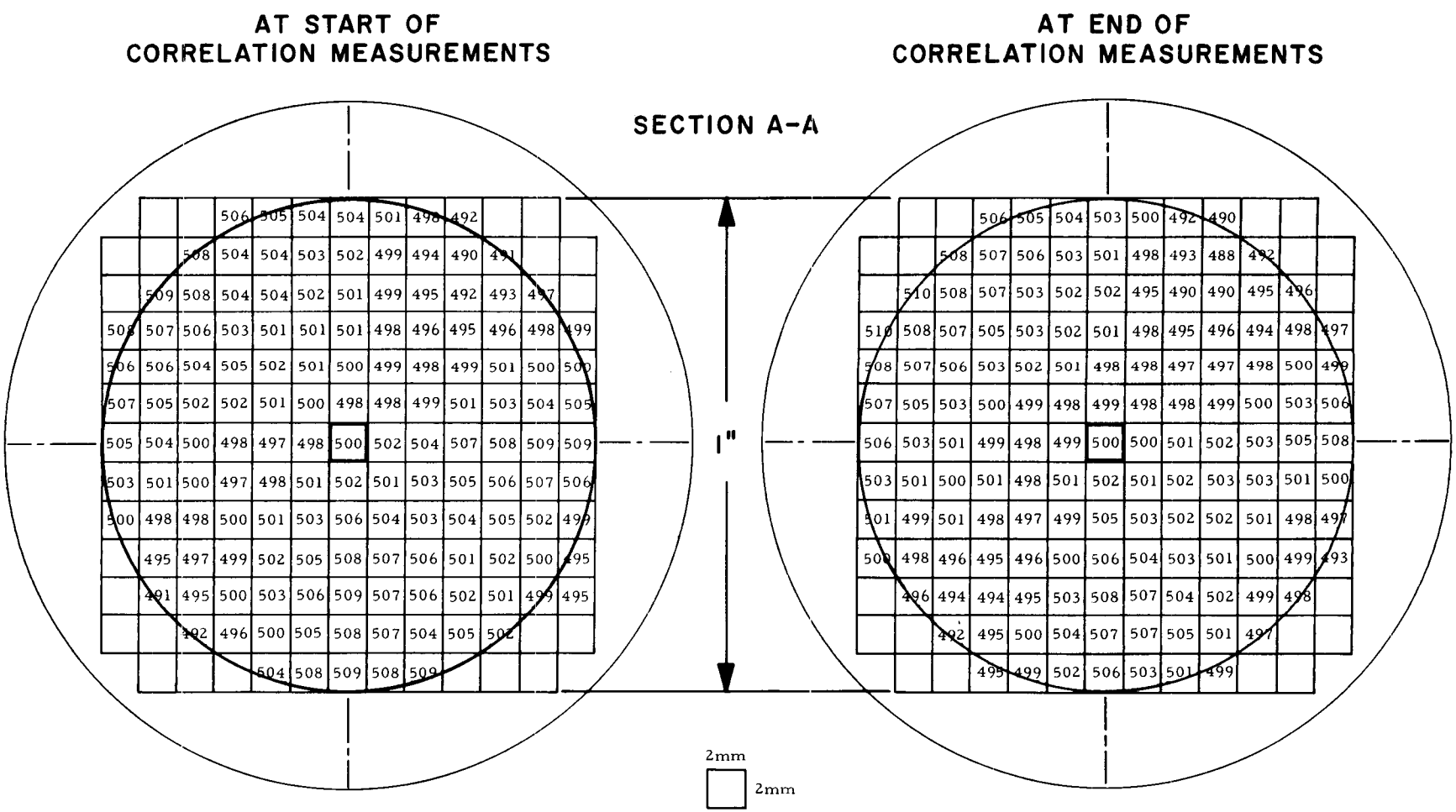


Figure 4-11. Relative Flux Densities of Collimated Beam at Position Shown in Figure 4-12.

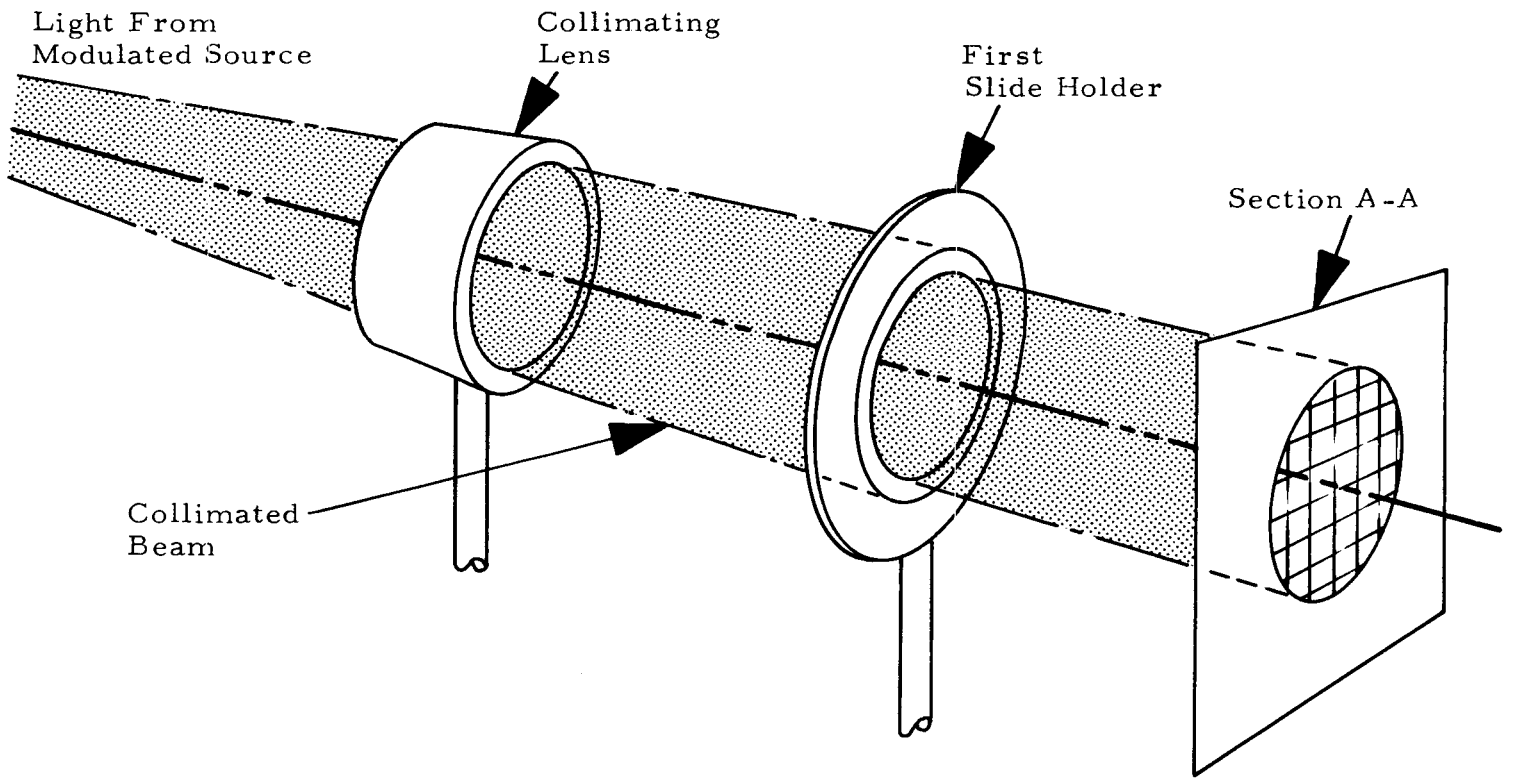


Figure 4-12. Configuration of Collimated Beam for Flux Density Measurements

to 240 cps and coupled to the microvolter as shown in Figure 4-13. The calibrated signal from the microvolter was coupled to the input of the two microvoltmeters in turn. The microvolter and audio oscillator were calibrated by the Standards Laboratory of Douglas Aircraft Co.

4.5.3.2 System Calibration

The system was designed so that an increase in the quantity of light transmitted through the slide holders would give a proportional reading on the microvoltmeter. If a series of metal discs with centered circular apertures of various sizes are inserted in the slide holder one at a time and the microvoltmeter reading recorded for each disc then the meter reading will be proportional to the area of the aperture being measured, and thus to the light transmitted through the slide holders, provided that the system is properly calibrated and linear. Figures 4-14 and 4-15 show calibration curves of the system for two settings of the main channel microvoltmeter gain control. Thus, by adjusting the gain control and shifting the scale readings, a range of 100 to 500,000 μv was obtained. During all of the calibration procedures the filament current remained constant.

All correlation measurements with two slides in tandem were made with the 0.005π area aperture measured as 5000 μv on the main channel microvoltmeter. For some pairs it was necessary to set the meter to the 2500 μv scale and then double the reading.

4.5.3.3 Check of Calibration During Measurements

At the start of each data recording period the following procedure was carried out.

The equipment was turned on and allowed to stabilize for 30 min. After stabilization of the equipment, previously recorded measurements were repeated and compared with the original readings. If any discrepancies were evident, the optical alignment of the system was checked and the components of the system were correctly positioned to duplicate the previous measurements. After the repeatability of the measurements had been established, correlation measurements were resumed. At the end of a measurement period the above procedure was

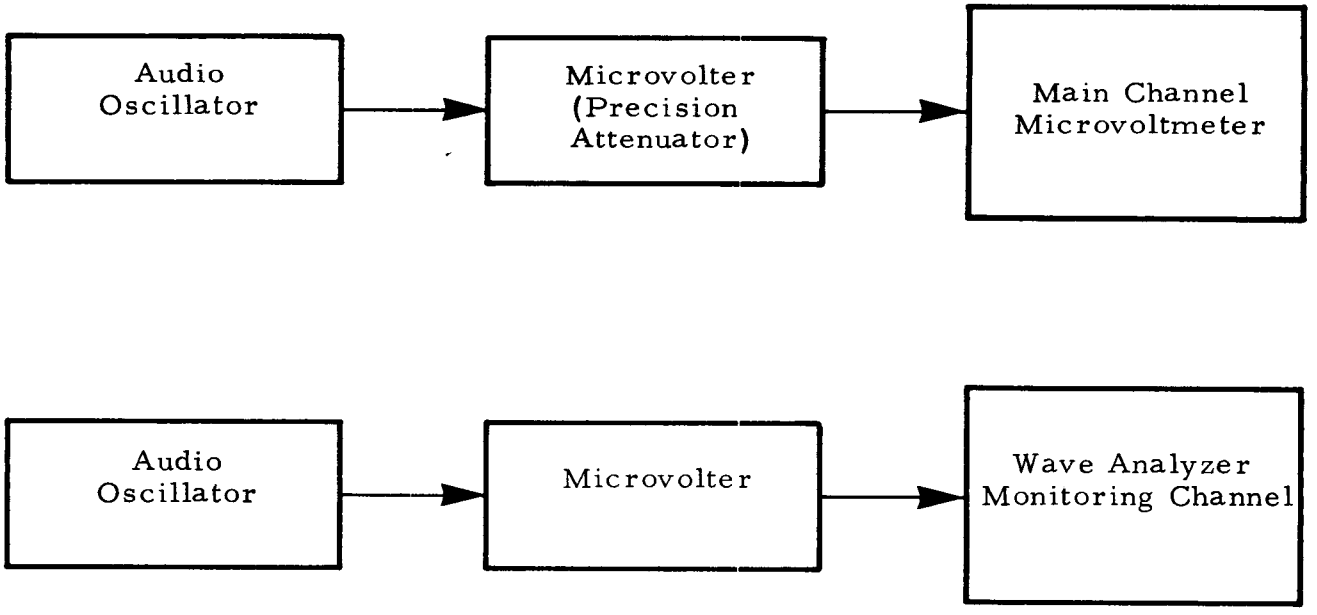


Figure 4-13. Calibration of Microvoltmeters

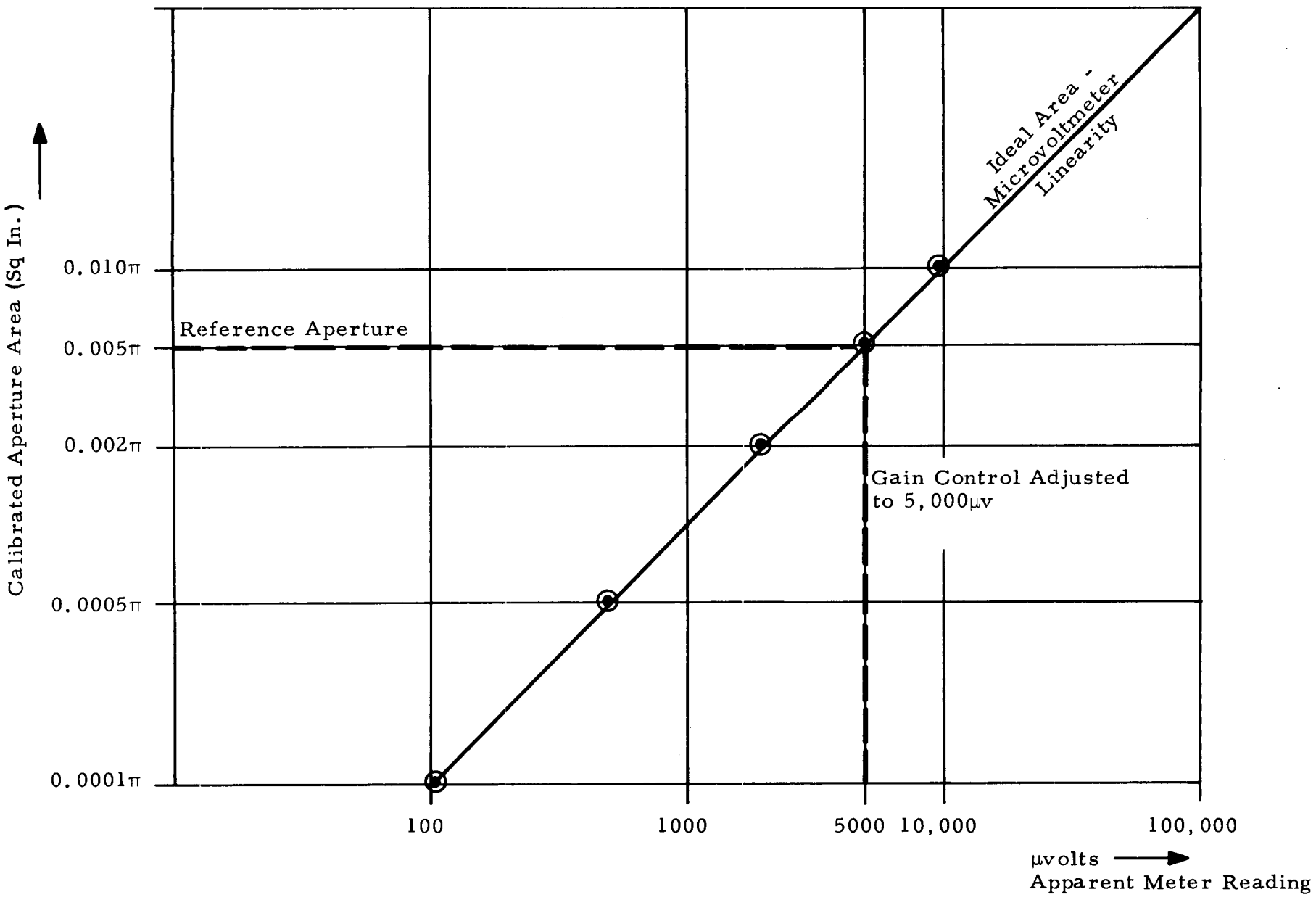


Figure 4-14. Linearity of Detector-Meter Combination (Gain Adjusted to 5,000μv for Aperture of 0.005πSq In.)

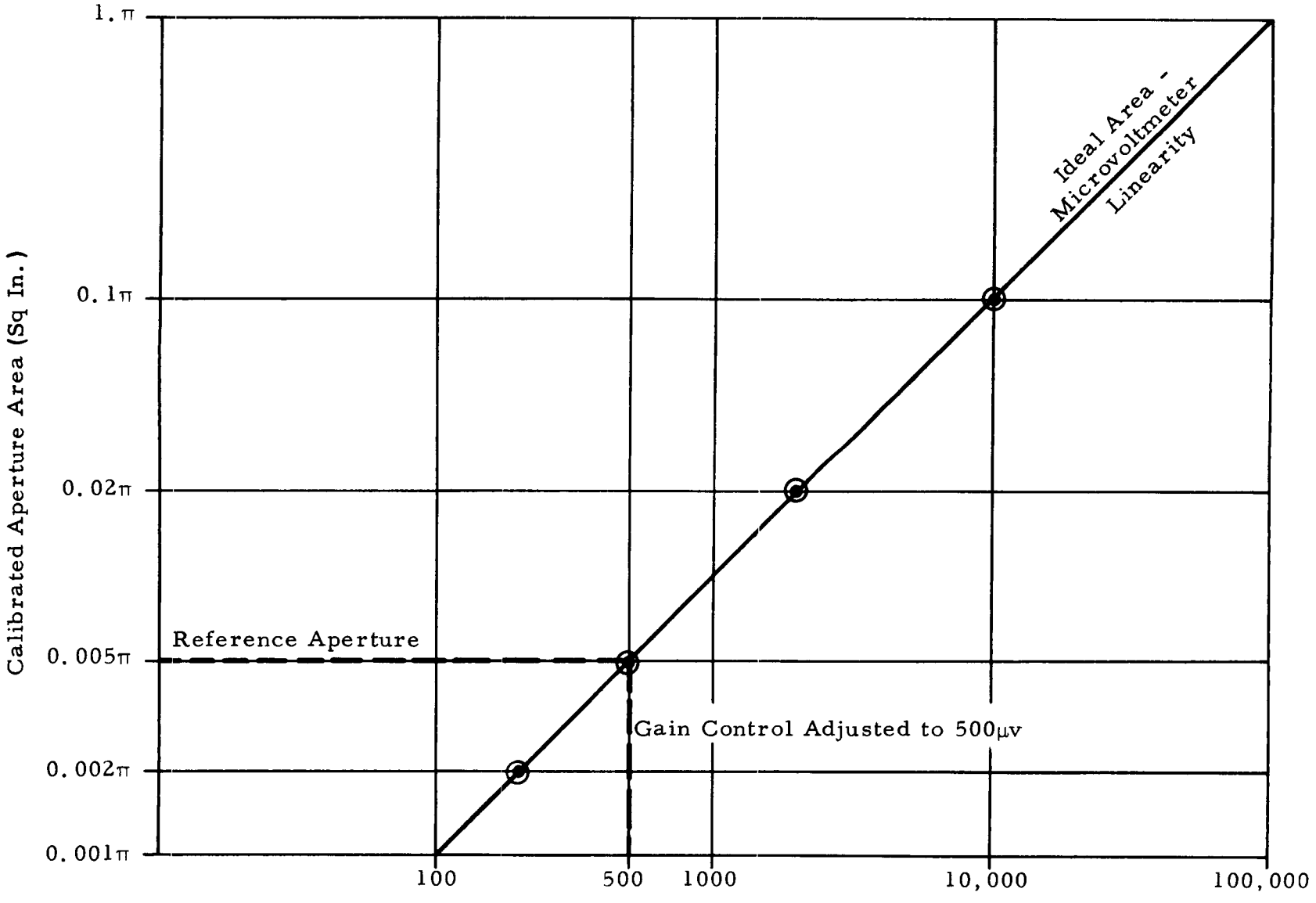


Figure 4-15. Linearity of Detector-Meter Combination (Gain Adjusted to 500μv for Aperture of 0.005π Sq In.)

repeated to ascertain that the equipment alignment had not been disturbed. During the measurements the output of the source was continuously monitored by means of an oscilloscope display. After every fourth set of measurements (a set of measurements being the six readings taken with a particular set of two slides), the calibrated reference disc was inserted in the second slide holder (the reading was not a function of which slide holder was used, since the light beam was well collimated) to check the calibration of the system. A scale deviation of more than 1% was uncommon, and if a scale deviation of more than 3% was indicated, the previous readings were checked.

4.5.4 Sequence of Correlation Measurements

4.5.4.1 Selection of Good Pairs

During the measurements of the first slides received*, the equipment was being modified to more nearly meet the system requirements for making the correlation measurements. The light beam as yet did not have the desired degree of flux density and intensity stabilization.

Although the uniformity achieved at this time was only about 15% (maximum variation from point to point in the A-A field of Figure 4-12), the flux density at the various points remained constant with respect to time. Thus valid transmittance measurements of identical slides could be made to obtain useful slide pairs ("useful" being defined by the 5% criterion - see Section 4.3). Since the value of the transmittance of the various slide sets was not required at this time, the determination of the useful pair was valid as long as each slide of any particular set being measured was identically positioned in the light beam. Initial difficulties with the instability of the light source were temporarily overcome by careful monitoring of the intensity and by repeated checks of the measurements. Although the latter was time-consuming, it did afford a method of checking slide sets for useful pairs while the equipment was being perfected, and it did minimize the delay due to the time required for producing new sets of slides to replace those sets from which useful pairs had not yet been obtained. After obtaining

*The slides were fabricated in lots as described in Section 4.3.

useful pairs for all 100 sets, the final development and calibration of the system was completed, and the useful pairs were again measured and referenced to the final calibration.

Three measurements were made on each of the selected slide pairs. Each slide of the pair was in turn positioned in the second slide holder with the slide index mark at 0°, and the transmittance through the single slide was measured. Following the two single-slide measurements, the higher reading slide was placed in holder #2 and the lower reading slide in holder #1. The transmittance through both slides was then measured with the index marks both set on 0°.

4.5.4.2 Routine Correlation Measurements

The bulk of the measurements were of the transmittance of two slides from different sets mounted in tandem in the light beam.* One slide was placed in slide holder #2 and the other slide was placed in slide holder #1, which was capable of 360° rotation in the plane of the slide. The slide holder was engraved with degree graduation so that the slide could be accurately positioned within $1/2^{\circ}$ of any designated rotational position.

Six measurements were made for each pair of slides inserted into the holders. For all six measurements, the slide in holder #2 (closest to the detector) was inserted and fixed with its index mark at 0°. For the first measurement, the slide in holder #1 was positioned so that its index mark was at 0°. The slide was then rotated 60° clockwise (as viewed from the detector) for each of the remaining five measurements. Upon completion of the six measurements, the pair of slides was replaced by the next pair, as designated by the logbook. This procedure was followed to complete the required 1600 sets of measurements.

* Of each useful pair, only the slide with the lower transparency was used for the cross-correlation measurement.

5.0 THE COMPUTER INVESTIGATION

5.1 System Parameters Under Investigation

The system under investigation is specified to be a forced learning perceptron. The logic units in the perceptron are specified to have their input connections arising from sensory points selected at random according to a uniform distribution. A representative set of patterns was selected. The remaining system parameters are:

a. Logic unit parameters

b. Number of logic units

c. Output unit threshold

The computer investigation may be characterized as a search for a suitable set of logic unit parameters. The optimum threshold for the output unit and the number of logic units required to approach asymptotic performance were obtained for the various logic unit parameter sets.

The logic unit parameters of interest are the number of input connections per logic unit, the weights assigned to the connections, and the logic unit threshold. The logic unit parameters are completely described by the two numbers:

$$\alpha = \frac{\Sigma w_i}{\sqrt{\Sigma w_i^2}}$$

$$\beta = \frac{\theta}{\sqrt{\Sigma w_i^2}}$$

where the w_i's are the weights assigned to the connections (and may be positive or negative) and θ is the unit threshold.

5.2 General Form of Program and Sequence of Computer Runs

To perform the estimation procedure, the equations of Section 3.1 were mechanized on an IBM 7094. The computer facilities of Douglas Aircraft Co. at Santa Monica, Calif. were utilized. The program was written in FORTRAN II. A copy of the compiled program and the FORTRAN program are included as Addendum A to this report.

The routine consists of a main program and two subroutines. Subroutine MEMR computes the empirical distribution of Q_{jt} from the equations with limits imposed by the patterns being tested. Subroutine OUT merely provides a means for printing out some of the results in a suitable format.

The 9600 data points (1600 cards) were processed for a number of parameter combinations. The purpose of these repeated trials was to attempt to find an optimum for the input connections to the logic units and a logic unit threshold, one that would process the data with a minimum of errors. Each computer run consisted of seven parameter combinations and took approximately 10 min of computer time.

In order to permit conducting a more extensive investigation, the possibility of using a fraction of the available data was investigated. The program checkout decks involving 120, 720, and 840 optical correlation measurements were completely inadequate. A set of 2400 measurements was carefully selected to provide 24 measurements on Q_{k+} and 24 measurements on Q_{k-} for each value of the index k. The results obtained with these short data decks were compared with those obtained using the full set of 9600 correlations (96 observations on each Q_{k+} and Q_{k-}). The means and standard deviations returned by the short deck generally differed by 20 to 50% from the values given by the full deck. This accuracy was considered inadequate, and the use of partial data decks was discontinued.

5.3 Description of Program

The implementation of the program consisted of the following:

a. Read in half of C table (integral of the normal distribution), compute other half of C table. Compute μ^2 and $\frac{e^{-\frac{\mu}{2}^2}}{\sqrt{2\pi}}$.

b. Compute $D_1 = n_x - n_y$ and $D_2 = \sqrt{n_x + n_y}$, multiplying factors for the mean and standard deviation of the patterns of one input to convert to many inputs.

c. Read in the single slide (T_{1j} and T_{2j}) and the single slide squared (T_{jj}) measurements and normalized by ($1/T_o$).

d. Read in remaining optical correlation measurements (T_{jt}), six rotations per slide pair, and normalized by ($1/T_o$).

NOTE: Since storage restraints did not permit operating on all data points during one pass through the computer, the correlations were performed in groups of 20 data cards per pass (loop 83).

e. Determine the class of the pattern under consideration (from its index number).

f. Compute the correlation (ρ) between the two patterns being processed for each rotation and $w = \sqrt{1-\rho^2}$.

g. Compute the limits of the integral.

h. Compute Q_j and Q_t (single-slide probability of activating a logic unit).

i. Check the limits to determine if they are above or below the bounds set on the integral, or equal to zero. Check integral multiplying factors $\neq 0$.

j. Compute the empirical distribution of Q_{jt} where

$$Q_{jt} = \frac{be^{-\frac{b^2}{2}}}{2\pi} \int_{\frac{a-b}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{\mu^2}{2}}}{\mu^2+b^2}\,du + \frac{ae^{-\frac{a^2}{2}}}{2\pi} \int_{\frac{b-a}{\sqrt{1-\rho^2}}}^{\infty} \frac{e^{-\frac{\mu^2}{2}}}{\mu^2+a^2}\,du$$

or modifications thereof, depending on the values of a and b.

k. Compute the average of Q_{jt} (Q_{k+} and Q_{k-}) for patterns in each class.

l. Compute the mean and the standard deviation for both pattern classes. Compute the standard deviation as a function of both the selection of the logic units and the distribution of the patterns.

m. Compute the difference of the means.

n. Compute the separation of pattern classes as a function of the number of logic units in the machine.

o. Select threshold which minimizes the number of errors on the asymptotic performance.

The computer routine returned the following outputs:

a. The expected values of the input to the decision element for patterns of the positive class and for patterns of the negative class, and the difference between these values

b. The inherent standard deviation within each class

c. The standard deviation for each class due to logic unit selection

d. The asymptotic class separation in standard deviations (see Figure 5-4)

e. The minimum number of misclassified patterns, asymptotic performance (see Figure 5-3)

f. A table of the expected input to the decision element as a function of the pattern

g. A table of the class separation in standard deviations, as a function of the number of logic units (see Figures 5-5 and 5-6)

h. The decision unit threshold producing minimum errors in asymptotic performance, and a table of these errors

i. A table of the probability of a logic unit being active, as a function of the pattern

j. Tables of Q_{k+} and Q_{k-}

k. Tables of the empirical distribution of Q_{jk} for two vortex patterns, one vortex and one nonvortex pattern, and two nonvortex patterns

A simplified flow chart of the computer routine is shown in Figure 5-1. Figure 5-2 is a detailed flow chart. Table 5-1 defines the symbols used in the computer program. The computer printout for seven of the parameter combinations is included in this report as part of Addendum A. Addendum A also contains a copy of the printout of the 9600 optical correlation measurements used as the test data in this project. Also included is a binary deck of the entire program and the test data on IBM cards.

5.4 Results of Computer Program

Using the final program and the full data set, some 77 combinations of α and β were investigated. In addition, 36 other combinations were investigated (16 with the full data deck) with an earlier version of the program. The earlier program did not provide all of the outputs desired, but was adequate to determine the suitability of the parameters.

The 77 points were obtained in three passes on the computer. In the first pass, 28 scattered points were selected, the choice being guided by the earlier partial results on 36 points. Based in these 28 combinations, 28 more points were selected to provide a finer grid in certain areas. The third pass of 21 points examined the region around $\alpha = 0.9$ and $\beta = 0.0$, where the number of errors appeared to be approaching a minimum. Some 42 additional parameter combinations were deleted from the three passes by operator errors.

None of the parameter combinations resulted in good separation of the pattern classes. The approximation used in estimating the component of variance due to logic unit selection requires good class separation for high accuracy. In many cases, separation was so poor that negative estimates of this component of variances were returned. In some cases, the mean value of the input to the response unit for the positive class was less than that for the negative class. Two criteria were used in deciding whether a parameter combination warranted further consideration:

a. The mean value for the positive class had to be greater than the mean value for the negative class.

b. The variance estimates had to be positive.

TABLE 3-1

Q-TABLE FOR HORIZONTAL (+) AND VERTICAL (-) BARS ON 20 x 20 RETINA

Number of Input Connections Per Logic Unit with Weight of		$\Sigma\tilde{Q}_{++}$	$\Sigma\tilde{Q}_{+-}$	$\tilde{Q}_{+}$	$\frac{1}{N}(E/\sigma)^2$		Exact Approx Error Factor
+1	-1						
X	Y				Approx	Exact	
3	3	0.14892	0.06340	0.05632	0.02269	0.02592	1.1
3	7	0.05408	0.01220	0.02475	0.01564	0.01523	1.0
4	4	0.25511	0.14600	0.08544	0.02324	0.02708	1.2
4	6	0.15396	0.06840	0.05848	0.02250	0.02473	1.1
5	5	0.37198	0.24740	0.11120	0.02237	0.02579	1.2

TABLE 3-2

Q-TABLE FOR ALTERNATE HORIZONTAL (+-) AND VERTICAL (+-) BARS ON 20 x 20 RETINA

Number of Input Connections Per Logic Unit with Weight of		$\Sigma\tilde{Q}_{++}$	$\Sigma\tilde{Q}_{+-}$	$\tilde{Q}_{+}$	$\frac{1}{N}(E/\sigma)^2$		Exact Approx Error Factor
+1	-1						
X	Y				Approx	Exact	
3	3	0.11408	0.09824	0.05632	0.00420	0.00578	1.4
3	7	0.03807	0.02821	0.02475	0.00368	0.00529	1.4
4	4	0.21222	0.18889	0.08544	0.00497	0.00646	1.3
4	6	0.12093	0.10143	0.05848	0.00513	0.00690	1.3
5	5	0.32457	0.29481	0.11120	0.00534	0.00664	1.2

TABLE 3-3

VALUES OF $1/N\,(E/\sigma)^2$ FOR THE ASTROPOWER DECISION FILTER

	Simulation Results	Use of Equation 11
$\frac{1}{N}(E/\sigma)^2$ for Sphere/Cube	0.150	0.145
$\frac{1}{N}(E/\sigma)^2$ for Pyramid/Ellipsoid	0.083	0.080

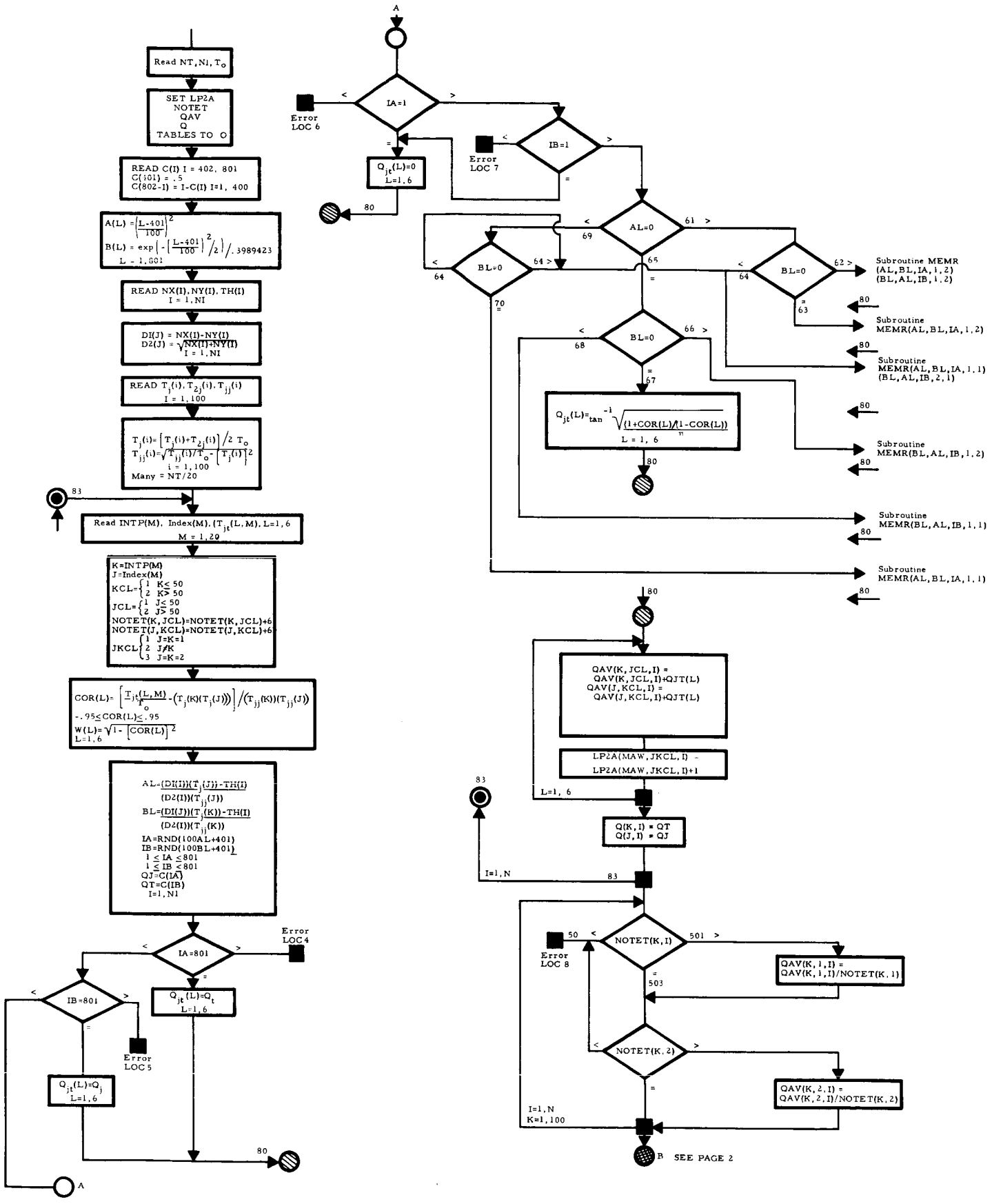


Figure 5-2. Detailed Computer Program Flow Chart

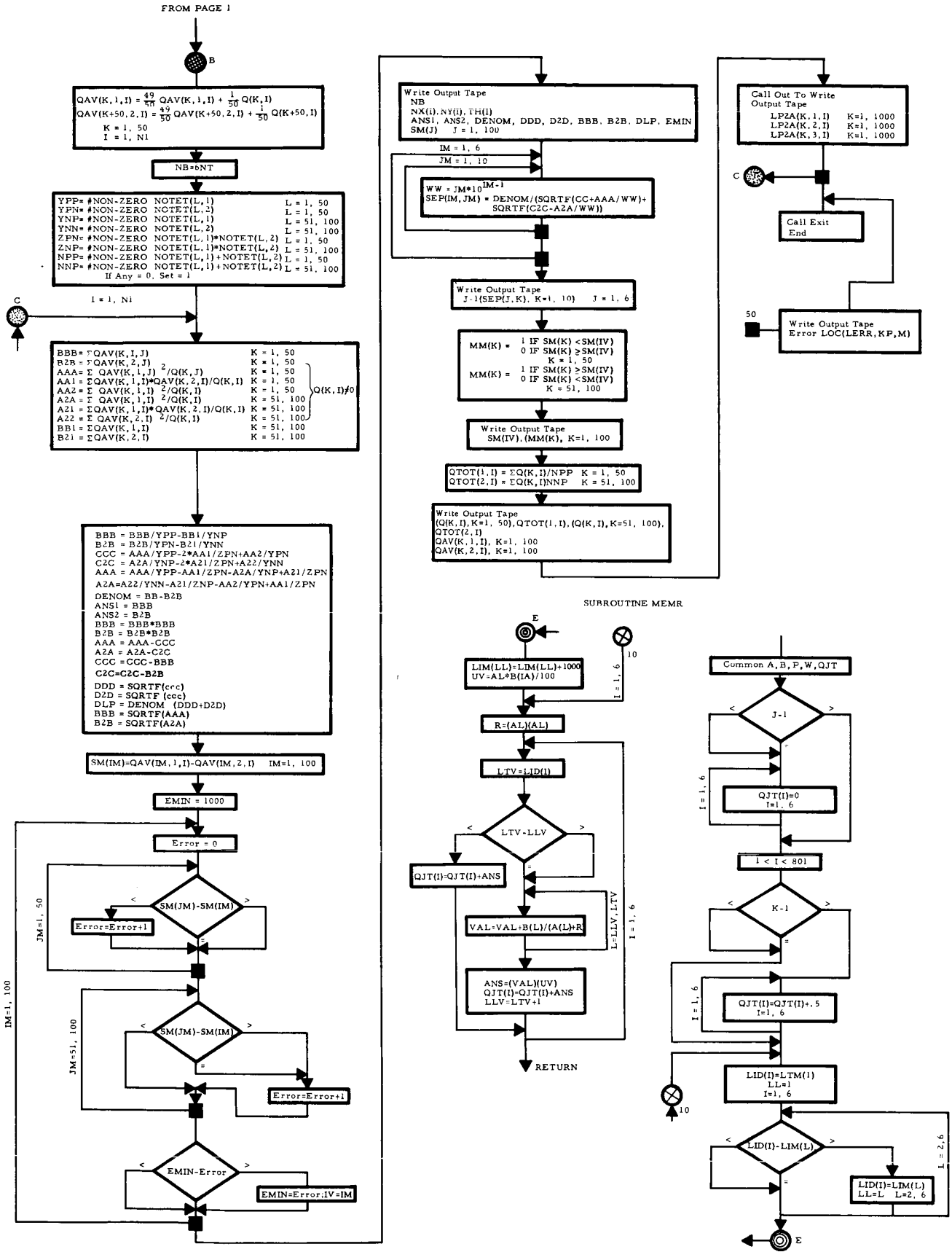


Figure 5-2 (Cont'd). Detailed Computer Program Flow Chart

TABLE 5-1

SYMBOLS FOR COMPUTER PROGRAM

A. Tables

The numbers in parentheses indicate the dimensions of the tables.

1. A(801) - squares
2. B(801) - normal density function
3. C(801) - normal integral
4. NX(7), NY(7), TH(7) - logic unit input parameters,

$$\alpha = \frac{NX - NY}{\sqrt{NX + NY}}, \qquad \beta = \frac{TH}{\sqrt{NX + NY}}$$

5. DI(7), D2(7) - α and β
6. TJ(100), T2J(100) - single pattern measurements
7. TJJ(100) - correlation of pattern with itself
8. TJT(6, 20) - two-pattern correlations
9. INDEX(20), INTP(20) - indices for two-pattern correlations
10. NOTET(100, 2) - number of samples for each Q_{k+} and Q_{k-} estimate
11. QAV(100, 2, 7) - Q_{k-} and Q_{k+} for each parameter combination
12. LP2A(1000, 3, 7) - empirical distribution for Q_{jt} for various vortex-nonvortex and parameter combinations
13. Q(100, 7) - probability of a logic unit being activated by the various patterns
14. QTOT(2, 7) - average of Q over the classes
15. SM(100) - expected inputs to the response unit for various patterns (inside loop on parameter combinations)
16. MM(100) - likelihood of error for given pattern, comparison of SM with best threshold (inside loop)
17. $\frac{COR(6)}{\sqrt{1 - \rho^2}}$, W(6) - inputs to subroutine equal to correlation coefficient ρ and respectively for two patterns in six rotations
18. QJT(6) - Q_{jt} returned from subroutine

B. Variables

1. NT, N1, TO - inputs defining number of two pattern pairs, parameter combinations, and reference light level for this pass
2. A21, A22, A2A, AA1, AA2, AAA, B21, B2B, BB1, BBB - used to accumulate partial sums of Q_{k+}, Q_{k-}, and products
3. YPP, YPN, YNP, YNN, ZPN, ZNP, NPP, NPN - counts on the number of entries in each partial sum
4. AAA, A2A, BBB, B2B, CCC, C2C, DDD, D2D, ANS1, ANS2, DENOM - variances, standard deviations, means, and their differences

TABLE 5-1 (CONT'D)

SYMBOLS FOR COMPUTER PROGRAM

5. AL, BL, IA, IB, AV, BV - used to obtain limits of integration
6. JCL, KCL, JKCL - indices denoting pattern classes
7. EMIN, ERROR - used in finding threshold for minimum errors
8. LERR, KP, M - indices used to specify error locations
9. NB - number of correlations this pass
10. QJ, QT - values of Q table for two patterns being tested

Only 11 of the 28 combinations in the first pass met these requirements. (It can be shown that failure to meet condition a. is an inadequacy of the optical correlation, and not of the perceptron technique. If good separation had been obtained, however, the measurement inaccuracies would not be of such significance.)

The most significant computer outputs are summarized in Figures 5-3 through 5-6. For each parameter set. a threshold for the output unit was selected to produce a minimum number of errors in the asymptotic performance. These minimum values are shown in Figure 5-3. The lowest value obtained was 32, showing the need for a closed-loop self-organizing process. A second threshold for the output unit was selected, this one providing maximum number of standard deviations for the threshold-to-class mean separation. These separations again are for the asymptotic performance. They are presented in Figure 5-4 and are seen to be about one-fourth as large as might be predicted from Figure 5-3. The absence of an entry on the chart indicates failure to meet requirement a. above. The number of logic units required to achieve separations equal to 99% and 99.9% of the asymptotic separation are given in Figure 5-5 and 5-6 respectively. Since the computer must return positive estimates of the variance for these calculations, Figures 5-5 and 5-6 reveal those parameter combinations for which the approximation failed to yield positive variances.

Summarizing the results, parameter combinations which could be analyzed at all were difficult to find. For those combinations which could be analyzed, class separations were poor, giving rise to high error rates. A network of about 40 to 50 logic units should provide essentially asymptotic performance for the better parameter sets. However, if the interclass separation were better, more logic units would be required to achieve asymptotic performance, and the error rate would be reduced significantly. At best a factor of 20 more separation seems desirable.

Limited experimental evidence (Tables 3-4 through 3-7) has indicated that more powerful self-organizing rules can provide substantial increases in separation compared with the random connection, forced learning

rule. In particular the iterative design closed-loop learning rule, together with discriminant analysis for selecting the logic unit input connections (Section 3.2.2) should provide sufficient power to achieve this separation of the pattern classes.

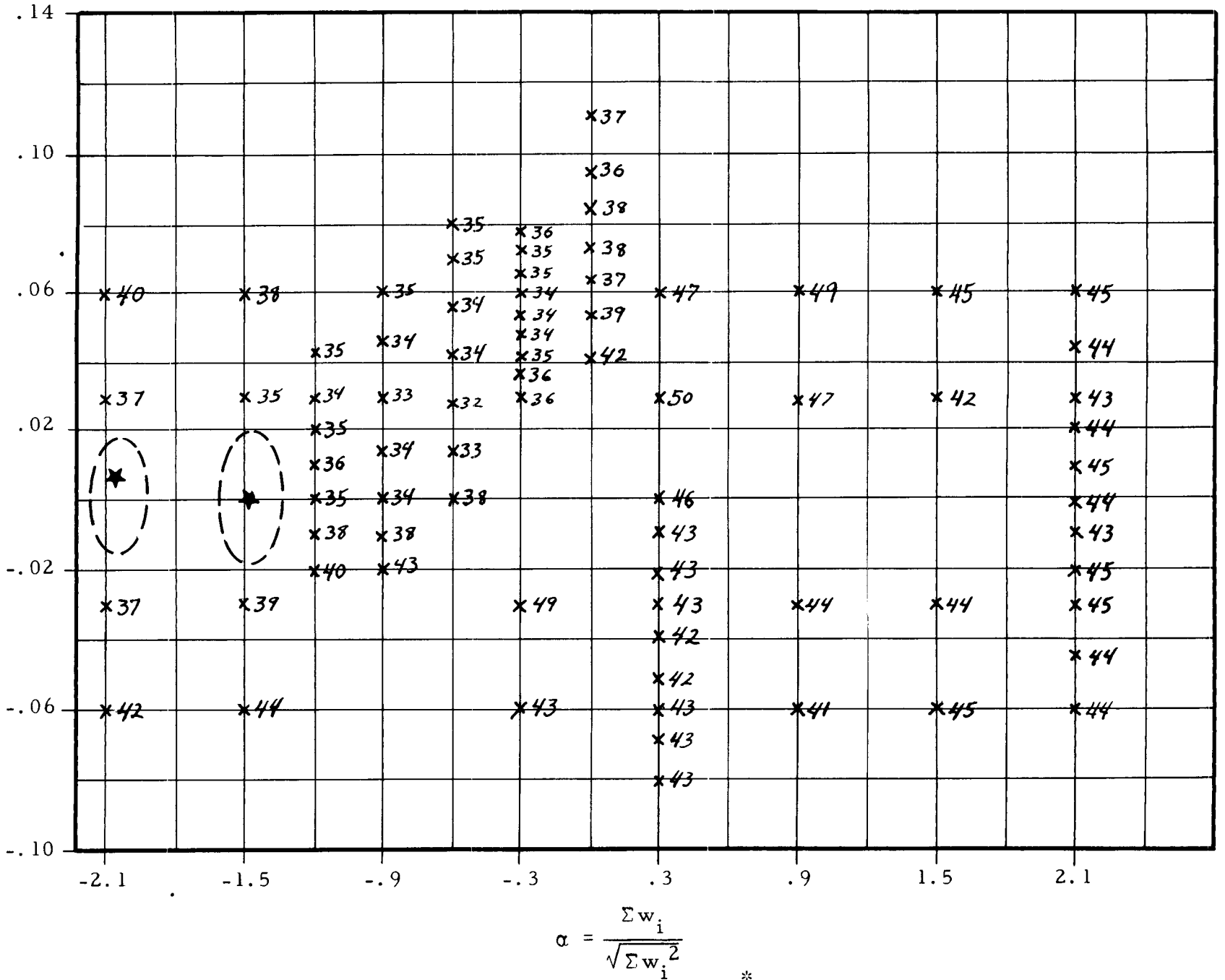


Figure 5-3. Minimum Number of Errors

*Data not available due to computer malfunction

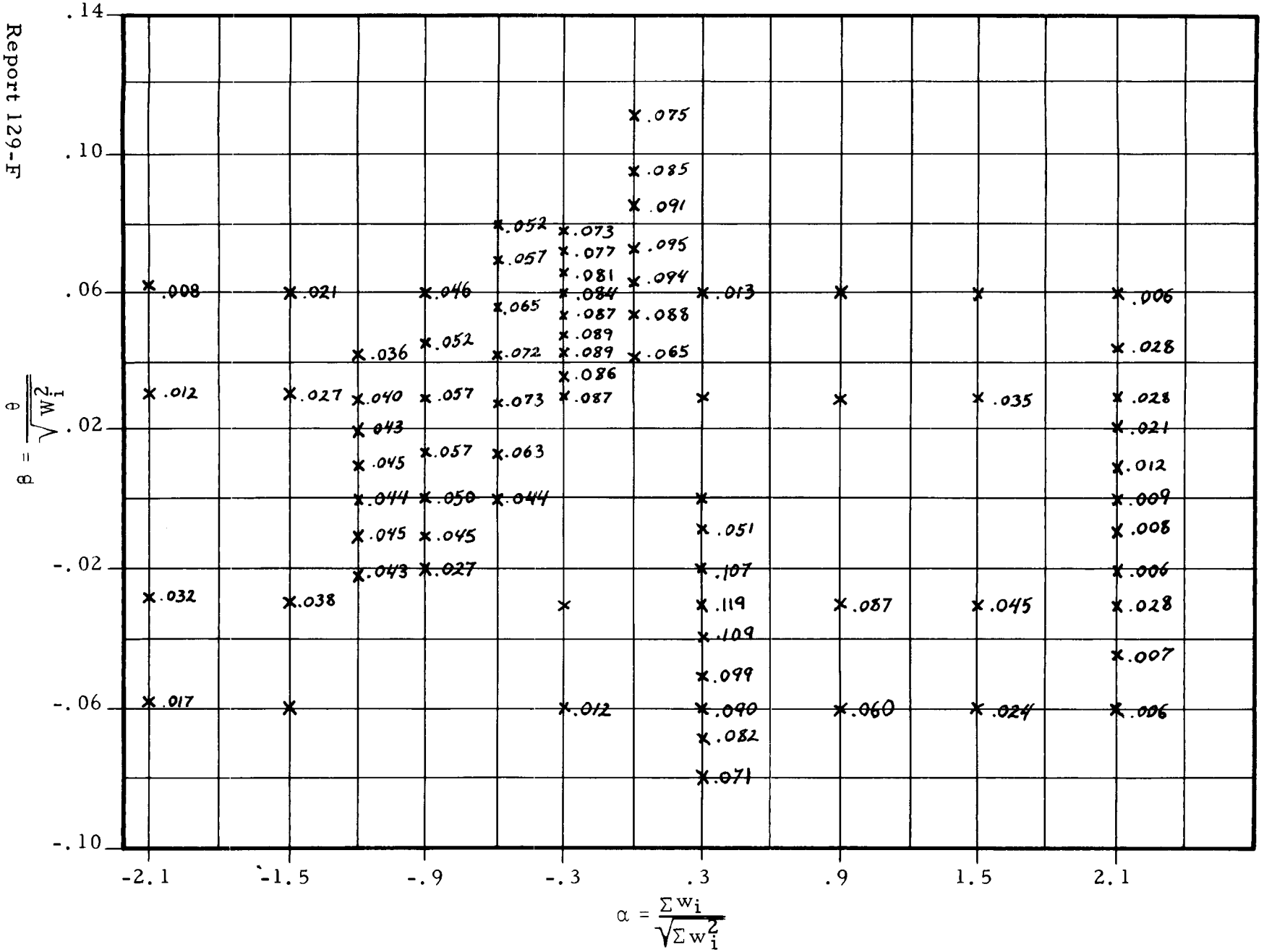


Figure 5-4. Asymptotic Class Separation in Standard Deviations

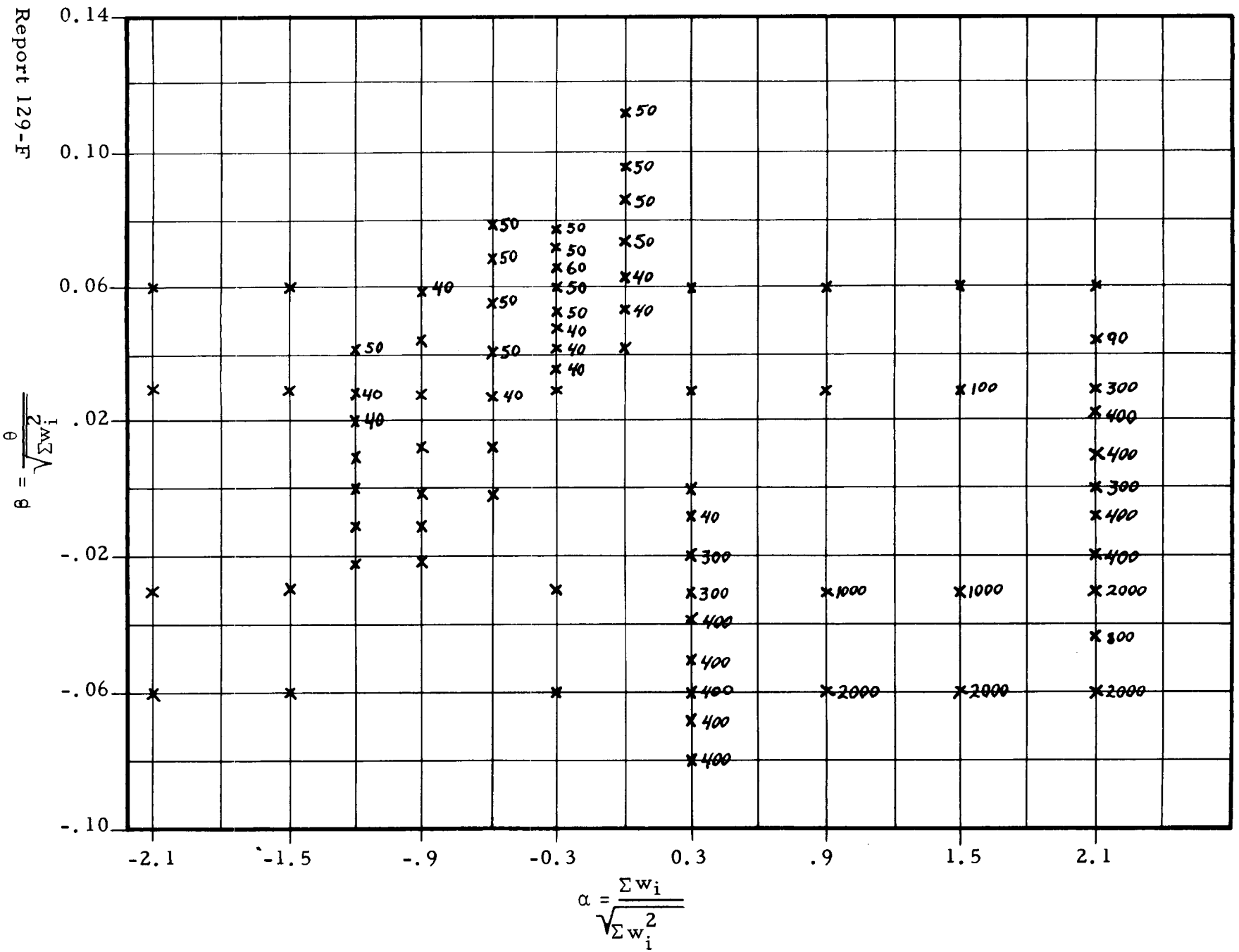


Figure 5-5. Number of Logic Units Required for 99% of Limiting Separation

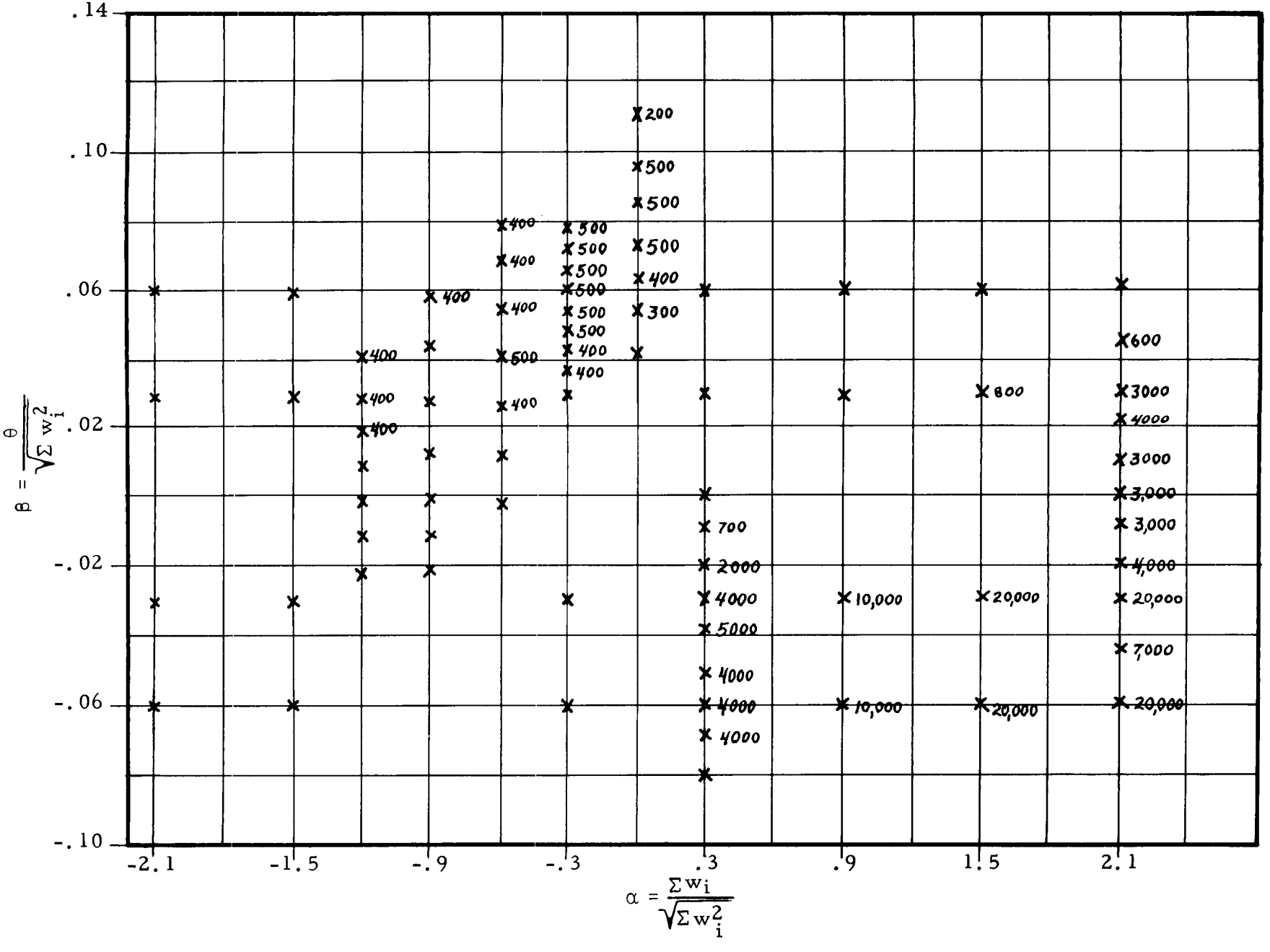


Figure 5-6. Number of Logic Units Required for 99.9% of Limiting Separation

6.0 HARDWARE REALIZATION

6.1 General Considerations

6.1.1 Available Data Format

Present TIROS satellites are equipped with a specially designed vidicon camera.* A solenoid-operated, focal plane shutter is used to expose the vidicon. The shutter speed is 1.5 msec. After exposure the viewed scene is stored on the screen. Readout is accomplished by a slow scan of 2 sec duration. The cloud cover data now has the form of a video signal of 62.5 kc bandwidth. This signal is transmitted via FM-FM to the ground station. There it is displayed on a cathode ray tube with a resolution of 500 lines per frame. A 35 mm camera is automatically triggered to produce a photographic negative of each scene displayed.

Future NIMBUS weather satellites will be equipped with three simultaneously operating cameras, one downward oriented and two oblique looking, to view a broad strip of the earth. Each slow scan camera will provide a resolution of 800 lines. Transmission to the ground station will generally be of the same nature as applied on TIROS.**

The primary input data to the envisioned recognition mechanism will in any case be a video signal, either directly from the vidicon scan or from a tape recording. A secondary form of the data will be the visual picture on the display scope, and a third possible form will be the photographic negative as used during this study.

As to the required processing speed, it suffices to know that NIMBUS will not transmit more than one 800-line picture per minute, which therefore allows a maximum of this much time for the recognition process. This would indicate a total delay of approximately 90 min from picture taking to completed interpretation.

---

* Two cameras are installed in TIROS satellites: a wide angle camera with an area coverage of 750 mi square, and a narrow angle camera covering 75 mi square. For this study, wide angle pictures have been used exclusively since they provide the area coverage required to view complete vortex structures.

** Schneebaum and Stampfl, "Data Storage for Meteorological Satellites," Astronautics and Aerospace Engineering, Apr 1963.

6.1.2 Required Picture Resolution

The study effort described in the previous sections was concerned with the number of logic units required to obtain satisfactory performance of a vortex recognition system. The number of logic units is only one parameter affecting the size of the recognition machine. Another as-yet-undetermined parameter is the number of sensory elements required to present the cloud data with adequate resolution to the logic unit layer.* The only way known at present to determine the required picture resolution (number of sensory points) is to try different resolutions. Certainly the number of picture elements will not be higher than the picture resolution of the original data, in this case 500 x 500 = 250,000 picture elements for TIROS data, and 800 x 800 = 640,000 picture elements for NIMBUS data. While for many of the large diameter vortex structures, such as Nos. 49 and 28 of Figure 6-1, a lesser resolution seems to suffice, some of the smaller diameter patterns, such as Nos. 36 and 55, seem to require (at least intuitively) the full available vidicon resolution.

In order to select a starting point for experiments, the following intuitive reasoning may be pursued. If a human can recognize a vortex pattern with, let us say, a resolution of 100 x 100 picture elements, then there is reason to believe that a machine could also achieve a satisfactory recognition. The number of 100 x 100 = 10,000 picture elements was arbitrarily selected, since it seems to be a realistic number of data points to be processed by machine per recognition cycle.

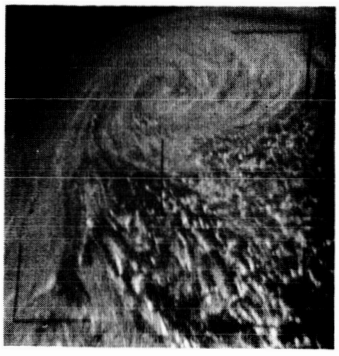
The photos of Figure 6-2 are presented to assist in this reasoning. These patterns are the same as the ones shown in Figure 6-1, but they were photographically processed to change either pattern size, resolution, contrast, or a combination of these parameters.

Photo 49 in Figure 6-2 has the same size of vortex as that shown in 49 of Figure 6-1. The resolution is reduced considerably by inserting a diffuser in the light passing between photo negative and enlarger base during the exposure. By looking at the same picture details in both figures, one can estimate the resulting resolution in No. 49 of Figure 6-2 as approximately 150 lines. The pattern can still be identified uniquely.

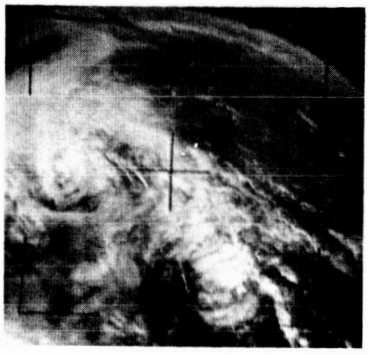
---

*Due to the approach taken in estimating the machine size with the aid of optical measurements, the number of sensory elements never entered the study as a parameter.

(49)

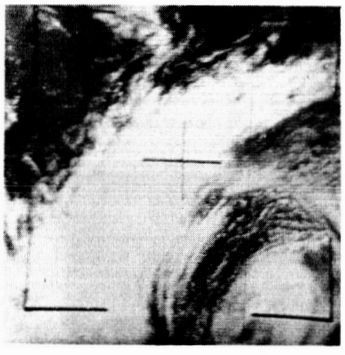


(50)

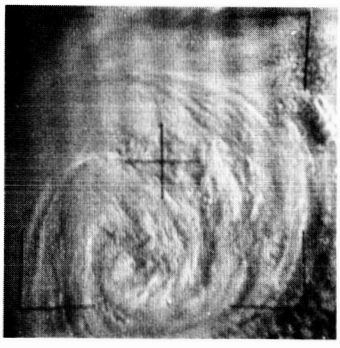


(36)

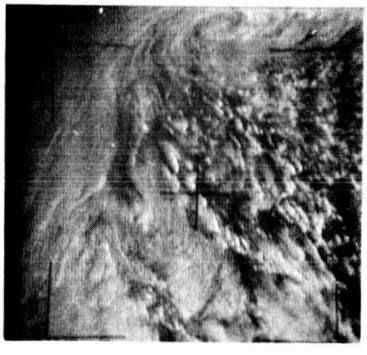
Variations in Vortex Size



(22)



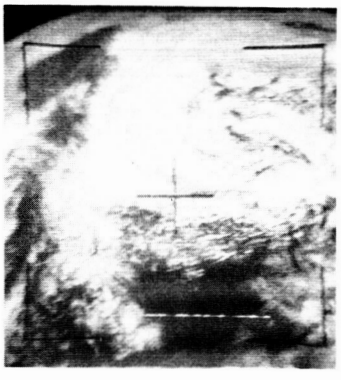
(28)



(31)

Variations in Picture Contrast

(45)

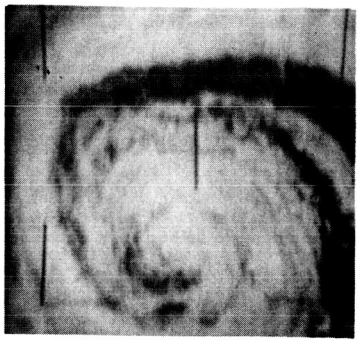


(18)

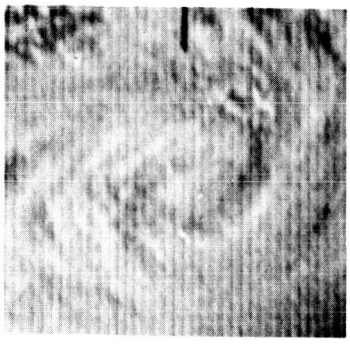
(55)

Differences in Pattern Definition

Figure 6-1. Examples of Cloud Patterns (Slide numbers indicated in parenthesis)



(49)

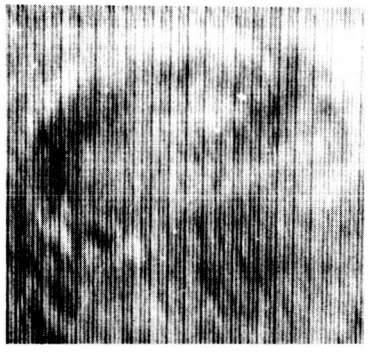


(50)

(36)

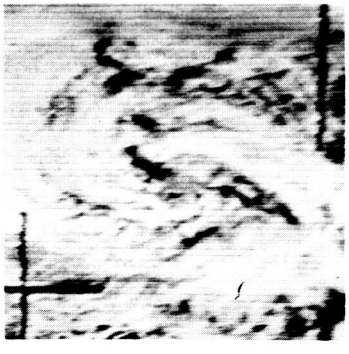
(22)

(28)

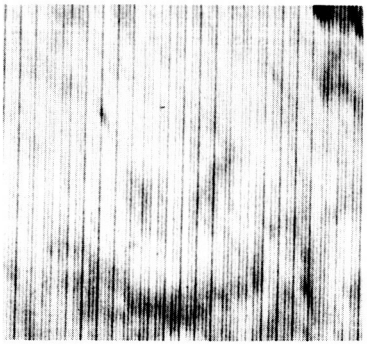


(31)

(45)



(18)



(55)

Figure 6-2. The Examples of Figure 6-1 After Photographic Preprocessing (Slide numbers indicated in parentheses)

Another example is photo No. 18 in Figure 6-2. It is an enlarged part of No. 18 in Figure 6-1, but it shows only the area containing the vortex pattern. The vidicon scan lines can be seen rather clearly, the resolution being approximately 200 lines. The vortex is still rather well defined.

Photo No. 31 in Figure 6-2 is an enlargement of the upper center area of No. 31 in Figure 6-1. The resolution is approximately 90 lines, and the pattern is still recognizable.

A closer inspection of Figure 6-1 also reveals two kinds of variations in the vortex pattern shown there. These are:

a. Variations in Vortex Size

The first row in Figure 6-1 demonstrates the size variation in the familiar vortex pattern. Cyclonic storms can reach a diameter of well over 1000 mi. The TIROS wide angle camera, when viewing vertically, covers an area approximately 750 mi on a side. NIMBUS will "see" approximately 1100 mi. Therefore, a vortex pattern can easily be spread over the complete picture area, as can be seen in photo No. 49. On the other hand, No. 36 shows a relatively small-sized pattern. Pattern sizes of 30 mi diameter can still constitute cyclonic storms of importance.

b. Variations in Picture Contrast

Depending on the sun angle, the cloud density and formation, and the angle of view, the light intensity and the contrast exhibited by the vortex pattern change considerably. The second row in Figure 6-1 shows this contrast variation. Photo No. 22 has very good contrast within the vortex structure, while No. 28, with the viewed area completely covered by clouds shows much less contrast. In No. 31, the vortex pattern is almost unnoticeable, while other areas of the picture can be recognized more distinctly.

Consideration of these two points and the previous remarks on resolution leads to the following scheme of preprocessing of the pictorial data. The complete picture is investigated for the presence of a large scale vortex with a resolution of 100 x 100 picture elements. Thus the recognition mechanism is presented with a 100 x 100 sensory field. In the second step, only one quarter of the picture is viewed, again with a resolution of 100 x 100, which now allows greater detail in a smaller area. As far as the recognition mechanism was concerned, it again works with a 100 x 100 sensory field, just as if it were being presented with a new picture. Considering overlap to assure that patterns are always viewed in their entirety, there are nine quarter-fields to be investigated in sequence, as shown in Figure 6-3. After this, 1/16 subfields can be investigated, these also with a 100 x 100 resolution, bringing the number of different readouts of one picture to the total of 59 shown in Figure 6-3. This preprocessing scheme would have these advantages:

a. The same recognition mechanism could be used for all picture areas under investigation, independent of area size.

b. Any vortex pattern would appear at least once in a rather standardized size; there would be no greater variation in size than 1:2.

The latter point could conceivably facilitate the recognition task considerably. This can be appreciated by comparing Figure 6-2 with Figure 6-1. In Figure 6-2 the pattern size is standardized to a certain degree, and recognition by the human observer seems to be more easily achieved.

Another preprocessing task that was simulated in producing the photographs of Figure 6-2 is a standardization of the mean gray level and the gray level variance. This conceivably could also facilitate the recognition task considerably, since it would reduce the contrast variations of the vortex patterns (see Figure 6-1).

It should again be pointed out that the number 100 x 100 is rather arbitrary and is presented in lieu of better information. Only experimental evidence can prove or disprove the adequacy of this particular resolution.

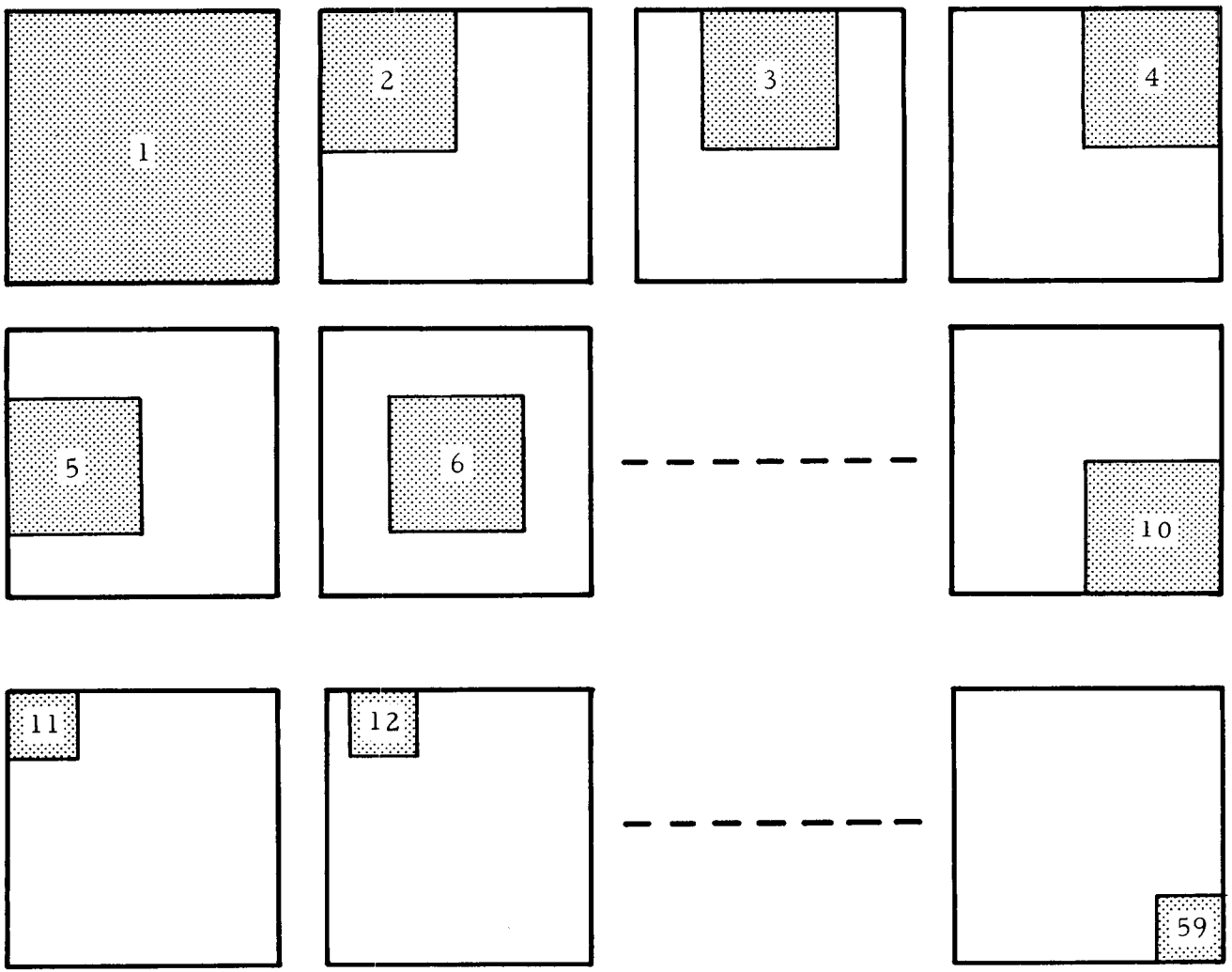


Figure 6-3. Readout Sequence During Preprocessing

6.1.3 Realization of Sensory Plane

The possible formats of the input data being specified, and a resolution of 100 x 100 picture elements tentatively proposed, there remains the question of how to present the data to the recognition mechanism. Since the proposed recognition mechanism is a parallel processor by nature, one first considers a parallel picture presentation. In this case the visual picture on a cathode ray tube would be the suitable input data format. The CRT display could be projected onto a 100 x 100 mosaic of photosensitive elements, such as photocells or photodiodes, or onto photosensitive material deposited directly on the tube face; or optical fibers could be used to transpose the individual picture. Each individual photo sensor would be followed by an amplifier to bring the power to a level suitable for the input to the logic units. This approach presents the following difficulties:

a. A 100 x 100 mosaic contains 10,000 photosensitive elements, and each amplifier will contain at least two transistors and a dozen other components. Thus the total number of components for the sensory plane alone will be well above 100,000.

b. The tolerances found in electrical performance of photosensitive elements and amplifier gain are such that an individual adjustment of 10,000 sensor-amplifier units will be required. Such an adjustment is not required on machines that operate on black-white inputs, as does the Astropower Decision Filter. In the case of cloud photographs, there are in the order of eight gray levels. This implies that all sensor-amplifier units must be equal in their input-output characteristic within at least 5%. An individual adjustment of 10,000 units is unfeasible.

c. The problems connected with the wiring of almost randomly distributed connections in such great numbers are severe. Also, once the wires are in place, modifications of the layout to improve performance are extremely cumbersome.

d. In order to investigate the total picture as shown in Figure 6-3 piece by piece, the following pre-processing sequence is required:

(1) Receive and store the overall picture with original resolution on a storage tube.

(2) Undestructively scan individual subareas as shown in Figure 6-3.

(3) Display the subareas on a second display tube (this tube is connected optically with the mosaic).

It is believed that this repeated picture conversion by analog means should be avoided if at all possible.

e. Many of the components required for this approach are not available off the shelf. Included in this category are items such as the photosensitive mosaic, the amplifiers, and the specialized scanning circuits for the first display tube.

A second approach is the serial digital presentation of the picture to the recognition mechanism. The original video signal is passed through an analog-to-digital converter, each picture element is represented by a binary word, and all words are stored in sequence in a magnetic memory. The reduction in resolution, the readout of the sensory point values into the recognition logic, the scanning of subareas, and the standardization of picture contrast are all achieved by digital computer techniques using a commercially available digital computer with associated peripheral equipment. Only comparatively little interface circuitry would be required. This approach is discussed in more detail in Section 6.2.

6.1.4 Realization of Logic Units

The performance of the logic units is determined by the equation

$$\text{Output} = 1, \text{ if } \Sigma \text{ inputs} \geq \text{threshold}$$

$$\text{Output} = 0, \text{ if } \Sigma \text{ inputs} < \text{threshold}$$

This characteristic can be obtained by any one of a number of conventional threshold circuits, such as the majority logic circuit used in the Astropower Decision Filter. For a large logic layer the number of components required, however, causes some difficulty.

The logic unit can also be instrumented in a digital computer fashion by location in the computer memory. This approach is very well suited if the sensory plane also has the form of memory locations, as discussed in 6.1.3. The logic manipulation can be accomplished in the arithmetic unit of a digital computer (see 7.2.2).

The choice of approach is very dependent on the implementation of the sensory plane.

6.1.5 Realization of Weights

The inputs of the logic units must be weighted, i.e., the electrical conductances of the connections that feed the unit inputs are varied in accordance with the learning rules. Several approaches have been tried to instrument such variable weights. Examples are motor-driven potentiometers,* multiaperture cores in connection with read-in and read-out circuits,** electrochemical resistors,*** and a combination of heating resistor and thermistor.**** These approaches, however, are cumbersome and uneconomical. Here again, digital computer techniques can be utilized. A location in memory is reserved for each weight. The content of a memory location can easily be modified according to any desired learning rule.

6.1.6 Rationale for Proposing a Digital Computer for the Hardware Implementation

In summary, there are three important reasons for choosing a digital computer for simulating the recognition mechanism rather than actually constructing a parallel logic computer.

---

* Used in the Perceptron of Cornell University.

** Brain, A.E., "The Simulation of Neural Elements by Electrical Networks Based on Multiaperture Magnetic Cores," Proc. IRE, Jan 1961.

*** The "Mimister" of Stanford Research Institute.

**** Mazetti, P., et al., "Experimental Construction of an Element of Thinking Machines," Kybernetik, Apr 1962.

First, there is no off-the-shelf hardware available to construct the parallel machine. Many items would have to be especially developed, such as the special scanning circuitry for the picture storage tube; and other components are presently beyond the state of the art, such as the 100 x 100 photocell matrix in the required form. On the other hand, digital computers of high capacity and high speed are presently available. A program to achieve the desired task seems feasible, and modifications to the standard A/D peripheral equipment appear to be manageable.

Second, a special purpose parallel logic machine with fixed wiring cannot easily be modified. It seems, however, that modifications will be desired in such a recognition network to extend its capabilities to new patterns classes that do not fit already established classifications. In the case of the digital computer, the modification of the program should be comparatively easy.

Third, a parallel logic machine, trained and wired for the recognition of vortex structures, is of no use in any other task. A general purpose digital computer, on the other hand, can readily be used for other computations as well. It should be noted in this connection that the National Operational Meteorological Satellite System Data Processing Center will have an IBM 7094 and several other smaller computers installed at its site in Suitland, Md.*

The initial thought that a special parallel logic machine could be built small enough to be incorporated in the satellite instrumentation must be disregarded in the light of present technology. Even if unexpected breakthroughs in integrated and miniaturized sensors should occur, the learning phase would still have to be achieved first on a digital computer. Also, the performance of the design would most economically be tested on a digital computer, so no effort would be wasted by achieving a digital computer design.

The next section, therefore, discusses the conceptual implementation of the recognition apparatus on an IBM 7094.

6.2 Computer Hardware Feasibility

The detailed consideration of a computer program for the automatic classification of cloud pattern pictures is far beyond the scope of this work. The following remarks represent no more than a preliminary brush with the problem, which will be completely successful if an order-of-magnitude cost figure for

*Johnson, D. S., et al., "Nimbus Data in Operational Meteorology," Astronautics and Aerospace Engineering, Apr 1963.

such program operation is achieved. The considerations are based on an IBM 7094 system, and on the availability of satellite cloud cover data in digital format (8 levels = 3 bits) on magnetic tape.

6.2.1 Data Organization

A single NIMBUS cloud picture contains about two million bits of information arranged in a nominal 640,000 words. The preliminary decision of handling 10,000 picture elements per cycle (which number is well suited to 7094 core memory size) requires considerable rearrangement of this data.

The first step suggested in this process is the construction of picture representation with resolution reduced by a factor of 2. This requires that each new picture element value be formed by the summation of four original picture element values. This results in a nominal five-bit gray scale rather than a three-bit scale.

Utilizing the feature of simultaneous computation and data input, it appears that these additions could be performed at a speed limited primarily by the speed of the data input. With use of the IBM 7340 hypertape drive, 28,330 words/sec may be inputed to the machine. Therefore, some 22 sec per picture are required for this accumulation procedure.

To facilitate the overlapping of these pictures, it is suggested that the basic storage block for the resultant picture representation be 50 x 50, and that after the accumulation of eight such blocks (corresponding to 50 rows of the full matrix) they be stored on another tape. Due to the rearrangement of these blocks, and to the large amount of memory they require, it seems that this operation may not be effectively made simultaneous with the accumulation process. Therefore, some 5-1/2 sec per picture are required for output of this information. It may then be expected that some 28 sec will be used in the first resolution reduction operation.

The two remaining resolution reduction operations may be done with more time sharing, assuming that two hypertape drives are used for storage of the half-resolution representations, and that a third is available to store the quarter-resolution picture. Total time for these operations should be on the order of 5 sec.

The gray level normalization may be accomplished through the summation of (1) the picture element values, and (2) the picture element values squared. Calculations based on these quantities allow a linear transformation of the actual element values to obtain the normalized ones. Roughly 20,000 multiplications (five bit words) and 30,000 additions would be required for such an operation for a 10,000 point matrix. The computation would take on the order of 1 sec per matrix.

In summary, it appears that a well equipped IBM 7094 system could achieve a useful size and gray-scale normalization with an expenditure of about 60 sec.

6.2.2 Decision Mechanism

The decision program is envisioned as beginning with the transfer of four blocks of input data into the computer, which then occupy 10,000 words of core storage. Using two hypertape drives simultaneously, this should be accomplished in less than 3 sec.

The basic program is seen as one which is repeated with different weights and connection addresses being introduced from disc storage. Such an operation may be achieved through use of the indirect addressing features of the 7094. Disc storage is suggested because of its great capacity and since the branching choice of subroutines, explained in Section 3.2.3, is almost ideally suited to the memory system capabilities.

The rate of information transfer from an IBM 1301 disc memory to core storage is approximately 15,000 words/sec. Assuming that a logic unit is characterized by 12 connections, each with an associated weight, and that each logic unit has a threshold and an output weight, 1000 logic units require 26,000 words of storage. Therefore, the data transfer speeds are approximately 600 logic unit specifications/sec. Transfer of an assumed 2000 logic units (see Section 5.4) therefore would require about 3.4 sec.

The calculations of positions of patterns, accomplished on basis of perhaps 500 logic units per iteration, could be expected to require about 0.3 sec/1000 logic units. Consequently, operation of the computer on one pattern could be expected to require some 0.6 sec. Some reduction of this time might be expected through simultaneous computation and data input, but

this seems difficult to achieve due to uncertainty as to the next subprogram to be used at any point.

Since some 59 patterns must normally be investigated to rule out the occurrence of a vortex in the NIMBUS data, roughly 4-1/4 minutes are required to process one return for each 1000 linear discriminants required to achieve recognition.

7.0 CONCLUSIONS AND RECOMMENDATIONS

The primary purpose of this program was to show the feasibility of the application of self-organizing techniques to the design of a parallel logic system for the automatic interpretation of satellite cloud cover photographs. This was to be accomplished in two stages:

a. The estimation of machine size required by the application of a learning algorithm that had been thoroughly documented mathematically - the forced learning procedure. The estimate was to be based on the results of optical correlation of a number of actual TIROS photographs containing both storm and nonstorm components. Only gross storm features (vortex structures) were to be involved in this first estimation task.

b. The extrapolation of the results obtained from the application of this open-loop, nonselective, random property algorithm to more powerful algorithms, such as natural selection and nonrandom generation of the property list, to increase logic unit efficiency and reduce machine size.

Since it was believed that a self-organizing system using forced learning could be made to asymptotically approach perfect performance for a reocgnition task of this nature, economic feasibility then could be determined simply by estimating the required complexity of the logic layer (in terms of the number of logic units) to arrive at some preselected closeness of approach to perfect performance.

7.1 Program Results

Optical correlations on 100 cloud cover transparencies (50 vortex, 50 nonvortex) were performed with a number of rotations of each pattern, giving a total of 9600 data points. This data was examined by a computer routine which mechanized the estimation procedure. Results of this program may be summarized as follows:

a. The nature of the patterns to be classified was such that good separation between pattern classes (vortex and

nonvortex) was not obtainable. This was due in part to the inaccuracy of the optical correlation procedure and photographic processing (5% overall accuracy was achieved through diligent photograph selection and control of the measurement apparatus). However, the primary cause appeared to be the complexity of the patterns themselves.

b. Computer analysis of the resulting test data showed that due to the lack of good class separation, the limiting performance of a forced learning perceptron was in the order of 65 to 70% of perfect performance. This level of performance could be achieved with 40 to 50 logic units.

c. The study of alternative self-organizing routines has produced more powerful algorithms involving a closed-loop decision process which promises considerable improvement of system performance by incorporating a learning operation having essentially two stages. The first stage is the implementation of a nonrandom procedure for the generation of the property list (discriminant analysis). The second stage is the maximization of the class separation by a closed-loop learning process which concentrates on those patterns most difficult to classify (iterative design).

7.2 Recommended Future Work

It has been shown that a forced learning perceptron is capable of recognizing vortex structures with some level of effectiveness (65 to 70%) despite the inherent complexity of the patterns. To increase the effectiveness of the recognition network to obtain near-perfect performance, the design of the recognition device using the more powerful learning algorithms developed concurrently with this project should be undertaken on a digital simulation program.

The main objective of the recommended work would be the design specifications required for the construction of a feasibility model - the proposed second phase of the present program. To obtain this objective, the tasks outlined below are required:

a. Determination of suitable logic unit parameters for use with a closed-loop learning algorithm

b. Determination of the limiting performance obtainable as a function of machine size for a closed-loop algorithm

c. Determination of the effect on logic unit effectiveness and efficiency of the inclusion of either discriminant analysis or a simplified form of discriminant analysis for logic unit specification

d. Determination of the effect of the number of sample patterns in the learning population on terminal performance, both for these patterns and for a validation sample of patterns not included in the learning population

e. Preliminary design of an automatic vortex or cloud pattern interpreter

f. System requirement study for a complete automatic cloud pattern recognition system

The design effort should be implemented in a digital computer simulation program. The investigation of discriminant analysis requires that the input patterns be available on digital tape, while the iterative design procedure is greatly facilitated by the availability of the data in this form. This effort should provide sufficient data for the detailed specification of a feasibility model of the parallel logic interpreter for the recognition of cloud structures indicative of storms.